THE ROYAL SOCIETY

Four centuries of geomagnetic secular variation from historical records

By Andrew Jackson[1], Art R. T. Jonkers[2] and Matthew R. Walker[1]

[1]*School of Earth Sciences, Leeds University, Leeds LS2 9JT, UK*
[2]*Department of History, Vrije Universiteit, Amsterdam, The Netherlands*

We present a new model of the magnetic field at the core–mantle boundary for the interval 1590–1990. The model, called *gufm1*, is based on a massive new compilation of historical observations of the magnetic field. The greater part of the new dataset originates from unpublished observations taken by mariners engaged in merchant and naval shipping. Considerable attention is given to both correction of data for possible mislocation (originating from poor knowledge of longitude) and to proper allocation of error in the data. We adopt a stochastic model for uncorrected positional errors that properly accounts for the nature of the noise process based on a Brownian motion model. The variability of navigational errors as a function of the duration of the voyages that we have analysed is consistent with this model. For the period before 1800, more than 83 000 individual observations of magnetic declination were recorded at more than 64 000 locations; more than 8000 new observations are for the 17th century alone. The time-dependent field model that we construct from the dataset is parametrized spatially in terms of spherical harmonics and temporally in B-splines, using a total of 36 512 parameters. The model has improved the resolution of the core field, and represents the longest continuous model of the field available. However, full exploitation of the database may demand a new modelling methodology.

Keywords: Earth's core; geomagnetic secular variation; magnetic field; maritime history

1. Introduction

The Earth has possessed a magnetic field for more than 4 billion years, generated in the fluid core. Although considerable progress has been made in elucidating the generation process (summaries can be found in this issue), much remains to be understood, not least why the field appears to be so stable for long periods of time punctuated by occasional reversals. Since humans have been observing the magnetic field for the past thousand years or so, and geographically diverse observations are available for the last 500 years, it is likely that there is much to be learned from an analysis of the field morphology and evolution deduced from direct measurements. To this end, we aim to construct the most accurate model of the magnetic field to date from original observations.

The models we create are specifically designed to examine the field at the core–mantle boundary (CMB), and our results are presented as maps of the radial component of the magnetic field at the surface of the core. Fortunately, the mathematical

tools for doing this (Whaler & Gubbins 1981; Shure *et al.* 1982; Gubbins 1983; Gubbins & Bloxham 1985) are well developed. It is now commonplace for dynamical simulations of the dynamo to plot maps of the radial field at the CMB to aid comparison with observation.

The historical record is very short compared with the time-scales used in palaeomagnetic analyses, or, indeed, compared with the life of the magnetic field itself. It is our opinion that it is important to try to push the analysis of historical observations back in time as far as possible, and, secondly, to improve the fidelity of our knowledge of the field in the recent past. In this paper we focus on our efforts to address these issues.

In terms of methodology, our work builds on that of Bloxham *et al.* (1989) and Bloxham & Jackson (1992); we omit describing their work, noting that it was the first systematic analysis of the core magnetic field morphology through time, and that it was entirely based on published data sources. As described in Bloxham *et al.* (1989), virtually all of the previous work on modelling the historical magnetic field had been designed to describe the field at the Earth's surface. Such models are not designed to be, nor are they, optimal for describing the configuration of field at the CMB.

A review of the previous compilations of magnetic data that have been produced over time can be found in Barraclough (1982). The earliest catalogues—of Stevin (1599), Kircher (1641) and Wright (1657)—are deficient in that they contain undated observations. Around 1705, the French hydrographer Guillaume Delisle compiled some 10 000 observations (mostly of declination) in his notebooks, trying to establish regularity in secular acceleration; these were never published but still exist in the Archives Nationales in Paris. The next important compilation of magnetic data was made by Mountaine & Dodson (1757), who claimed to have based their tables of declination at different points on the Earth on over 50 000 original observations of the field. It was this claim that partly provided the stimulus for our work. This claim was treated with some scepticism by some (e.g. Bloxham 1985) on the grounds that such large numbers of observations almost certainly did not exist at that time, since such a number was more than an order of magnitude larger than the number of observations in published works. Although they did not reference their data sources in a modern manner, nor publish the actual observations, Mountaine & Dodson (1757) do explicitly state the sources they used for their data, which are very similar to those we have used. We are therefore confident that their claim for the number of data that they used is true, as our own data collection activities show that such numbers of data clearly exist. It is also the case that we have used only a tiny fraction of the data that reside in the India Office of the British Library (we have used 325 out of the estimated 1500 pre-19th century logs that are held there; however, we have processed all of the 17th century logs). The early work of Mountaine & Dodson (1757) should almost certainly receive more prominence than it does, representing probably the first large-scale attempt to describe the morphology of the field. Maps based on the data were subsequently produced.

The catalogues that were produced after the work of Mountaine & Dodson (1757) have been the basis for the data compilation described in the study of Bloxham *et al.* (1989), and will not be discussed here, save for the remarks regarding the catalogue of Sabine given in § 2 *b*.

The plan of the paper is as follows. General aspects of the data that contribute to the new model are described in § 2; a more detailed account will be forthcoming. The methodology that has been applied to the dataset is described in § 3, along with the deterministic corrections that are made to the data. The stochastic model for navigational errors, and assignments of other errors, are described in § 4. The results are discussed in a geophysical context in § 5; for results of this analysis applied to the history of navigation and science see Jonkers (2000).

2. Data

The major problem affecting early maritime observations is that of navigational imprecision. By far the most important, and potentially the most accurate, observation on board ship was that of latitude. Even for the very first long-distance voyages, latitude could be found quite accurately; its determination had been a long-established practice by land-based astronomers, and it was initially their instruments, quadrant and astrolabe, that were adapted for use at sea. A more practical device on deck was the cross-staff, a stick with separate transoms of different dimensions sliding along its length. The problem of having to look directly into the sun was obviated by the backstaff, invented by John Davis in 1594, working with cast shadow instead of direct sighting. Thus, the problem of determining latitude has been rather well solved for the last 400 years.

As is well known, the problem of determining longitude at sea was not adequately solved until the 1760s (and gradually implemented in merchant shipping from the 1780s); the story of the battle of John Harrison to gain acceptance of his timekeepers has been oft told and we will not dwell on the issue (see, for example, Andrewes 1996). Before the use of accurate chronometers, or the competing lunar method, longitude was found by the method known as *dead reckoning*. At the beginning of each new ship's day (at noon), the 'day's work' was performed, either with or without an observation of latitude. The calculation of the change in longitude was based on simple trigonometry, using an estimate of the distance run by the ship (based on assessment of the ship's speed, taking into account currents and leeway) and the ship's heading. The result of ill-founded estimates of speed, leeway and drift were able to produce longitudinal errors of up to several degrees after months of unchecked progress. Much of our effort has been directed towards properly accounting for, or statistically representing, the effects of this poor knowledge of longitude. We have made independent estimates of the accuracy of dead-reckoning estimates using estimated and measured latitudes (see § 4 *a*).

The measurements of the magnetic field that were made most commonly in historical times were the declination (or angle between true north and magnetic north), denoted D, and the inclination or dip (or angle between the horizontal and magnetic field vector), denoted I. After Gauss's development of a method for measuring absolute intensities in 1832, the horizontal intensity (H) and the total intensity (F) were observed. In more recent times, X, Y and Z, the northward, eastward and downward components of the magnetic field, respectively, are reported. Exact definitions can be found in Langel (1987).

Awareness of magnetic declination dates back to the first half of the 15th century, at least in Europe. These first measurements take the form of scattered land observations. More importantly for our purpose, the oceans have been traversed

by ships from many nations since the 16th century. In marked contrast to coastal navigation, negotiating an ocean relied on astronomical observation for a daily latitudinal fix (accurate using a backstaff to about ten minutes of arc; e.g. Andrewes (1996, p. 400)), and a compass to maintain course in the absence of landmarks and soundings to steer by. In addition, compass bearings were used in dead-reckoning calculations (see below) and to chart coastlines. All of these applications necessitated an adequate correction for local declination. Navigational practice thus frequently included magnetic observations, as is evinced in thousands of logbooks kept by captains, masters and mates. Although many of these manuscripts have since been lost, a sufficiently large number has been preserved to warrant a substantial sample to be extracted for geomagnetic modelling purposes. In the following, we will briefly outline the shipping companies we investigated, their routes, and some characteristics of the measurements.

(*a*) *Pre-19th century data sources*

Our archival research took place in Great Britain, France, the Netherlands, Denmark and Spain. Three main categories of shipping have been analysed: East-India Companies, naval expeditions, and smaller merchant organizations. The first category provided about twice as many readings as each of the other two, and represents the earliest and longest record. The English East India Company (EIC) obtained a royal charter in 1600, while its Dutch counterpart, the VOC (Vereenigde Oost-Indische Compagnie), followed two years later. Both eventually suffered financial ruin on account of war losses and an ever-growing administrative burden due to territorial management. Three French, state-sponsored 'Compagnie des Indes' (CDI) were launched in 1664 (by Colbert), 1717 (Law), and 1785 (Calonne). None proved commercially viable, and only the second left any data. An even smaller player was the Danish 'Asiatiske Kompagni', founded in 1732 after a pioneering trading expedition to Canton.

All of the above ventures had both standardized instruments and a highly developed infrastructure, which allowed the processing and communication of nautical information (in logbooks, charts and sailing directions). The same can be said to a lesser degree of these countries' navies. During the 17th century, their peace-time rule was mainly confined to patrolling the Channel, the North Atlantic and the Mediterranean. Eighteenth-century convoy duties, however, expanded their actions across the Atlantic to the Americas (especially the West Indies); destinations eventually came to include the East Indies as well. The Danes maintained a military presence along the coasts of Greenland, where they collected a series of magnetic dip measurements in 1736. Of particular interest for the present studies are scientific voyages performed under naval command. Exploration was usually accompanied by particular attention to instruments and the observations obtained therewith, resulting in dense, high-quality records of otherwise sparsely visited regions. We merely recall in passing the Dutch exploits in the Pacific by the Nassau fleet, Edmond Halley's two magnetic surveys, Anson's and Cook's circumnavigations, and French expeditions by Bougainville, Crozet, Laperouse and d'Entrecasteaux.

Of the merchant navies, Dutch slavers (foremost the 'Midddelburgsche Commercie Compagnie', or MCC) and the English 'Hudson's Bay Company' (trading in furs) deserve mention. Both were active in the 18th century; records of small shipping companies tend to be extremely scarce before that time, and, indeed, the largest Dutch

slaver company (the West Indian Company) left very few ship's logs, in contrast to the MCC. A triangular route was pursued by the MCC: from Europe down to the West-African coasts (to buy slaves), then across the ocean to South America and the Caribbean (to sell slaves and buy agricultural produce), and homeward bound along more northerly latitudes. This triangular trade, in combination with direct traffic by Europeans to the West Indies and North-American colonies, generated magnetic field observations that covered much of the North Atlantic, even though the average number of measurements taken per voyage was lower than on ships bound for Asia. Relatively few measurements were taken on each voyage along the high-latitude track towards Hudson's Bay, but the regularity of the ships' annual crossings, which continued well into the 19th century, actually generates excellent data coverage.

The data that we have amassed for the period to 1800 is shown in figure 1*a–f*. Note that even in the first 50 years of the 17th century (figure 1*a*) there were some Pacific crossings, exhibiting the technique of 'running down the latitude'. Many of these data originated in the compilation of Hutcheson & Gubbins (1990), which was largely based on the work of Van Bemmelen (1899). In contrast, the last 50 years of the 17th century (figure 1*b*) are somewhat devoid of Pacific crossings, save for the data near the Aleutians. Between 1700 and 1750 (figure 1*c*) there is a tremendous increase in traffic around Cape Horn, and the voyages of Middleton into Hudson's Bay can clearly be seen. The latter part of the 18th century witnessed the tremendous explosion in voyages of exploration, generating a very even geographical coverage of the world (figure 1*d*). Note that there are data stretching from the Bering Straits in the north almost to the edges of the Antarctic continent in the south. The distribution of inclinations for the 17th century, shown in figure 1*e*, has been greatly improved from that of Hutcheson & Gubbins (1990), whose data were entirely in the Northern Hemisphere, consisting of observations in Europe and the Barents Sea, Hudson's Bay, and a single observation at Cape Comorin, India. Our dataset has inclinations in South Africa, the Philippines, Japan and along the east coasts of North and South America.

In all, of 2398 logs processed, 1633 contained a grand total of 83 861 magnetic declination observations, gathered on 64 386 days. For the 17th century alone, the former compilation of 2715 measurements was boosted with 8282 new readings. In extremely rare instances, inclination was also measured on board; the existing set of 1721 inclinations has grown with an additional 266 measurements.

(*b*) *19th and 20th century data sources*

Our data-compilation work has focused on improving the collections of Bloxham (1985) and Jackson (1989). Bloxham's 19th century data came entirely from the compilations of Sabine (1868, 1872, 1875, 1877); Jackson's data came from a variety of sources, documented in Bloxham *et al.* (1989), as no previous catalogue for this period had been made. We have made a number of improvements to this database. We have added the 12 post-1790 voyages contained in Hansteen (1819), which had not been previously used. Sabine's compilations of data cover the period 1820–1870. They are based on both published accounts of expeditions, British and foreign, and manuscripts to which Sabine had access. Fortunately, the sources of data are all listed. It is perhaps inevitable that the compilation is incomplete, and we have compiled many data from this period that were omitted from Sabine. A more serious

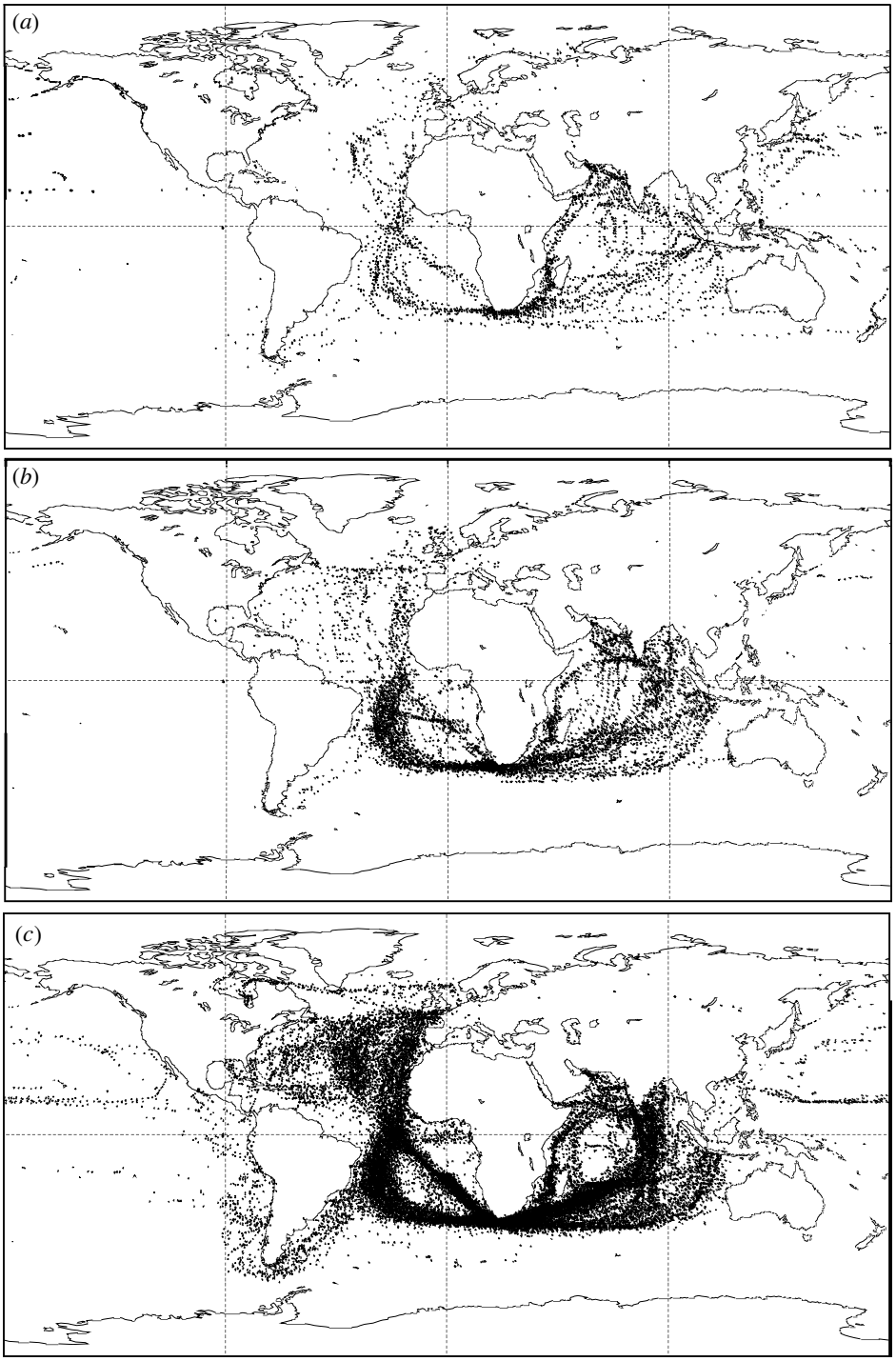

Figure 1. Data distributions in time. (*a*) Declinations 1600–1649. (*b*) Declinations 1650–1699. (*c*) Declinations 1700–1749.

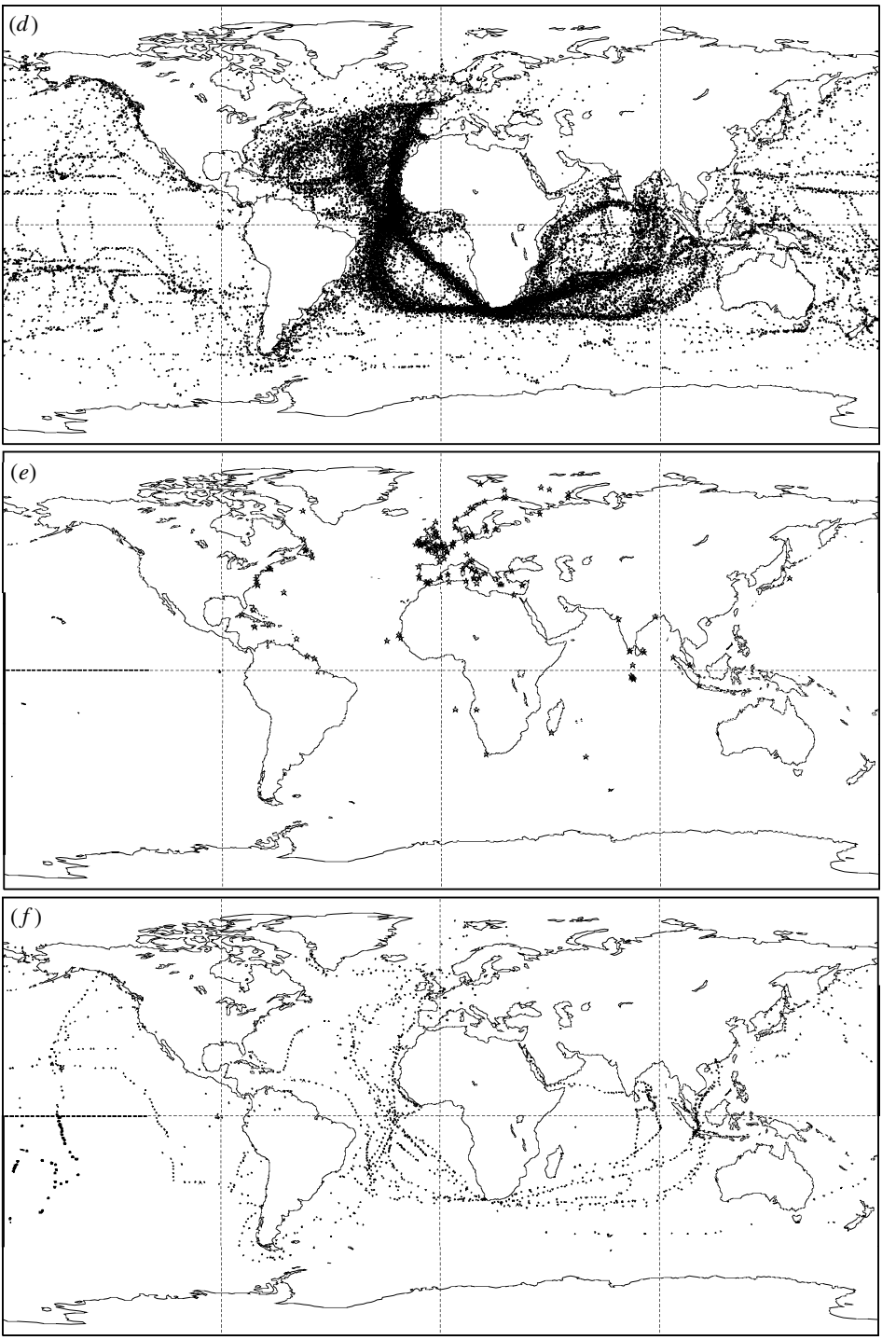

Figure 1. (*Cont.*) (*d*) Declinations 1750–1799. (*e*) Inclinations 1600–1699. Stars are used here to allow the few locations to be seen. (*f*) Inclinations 1700–1799.

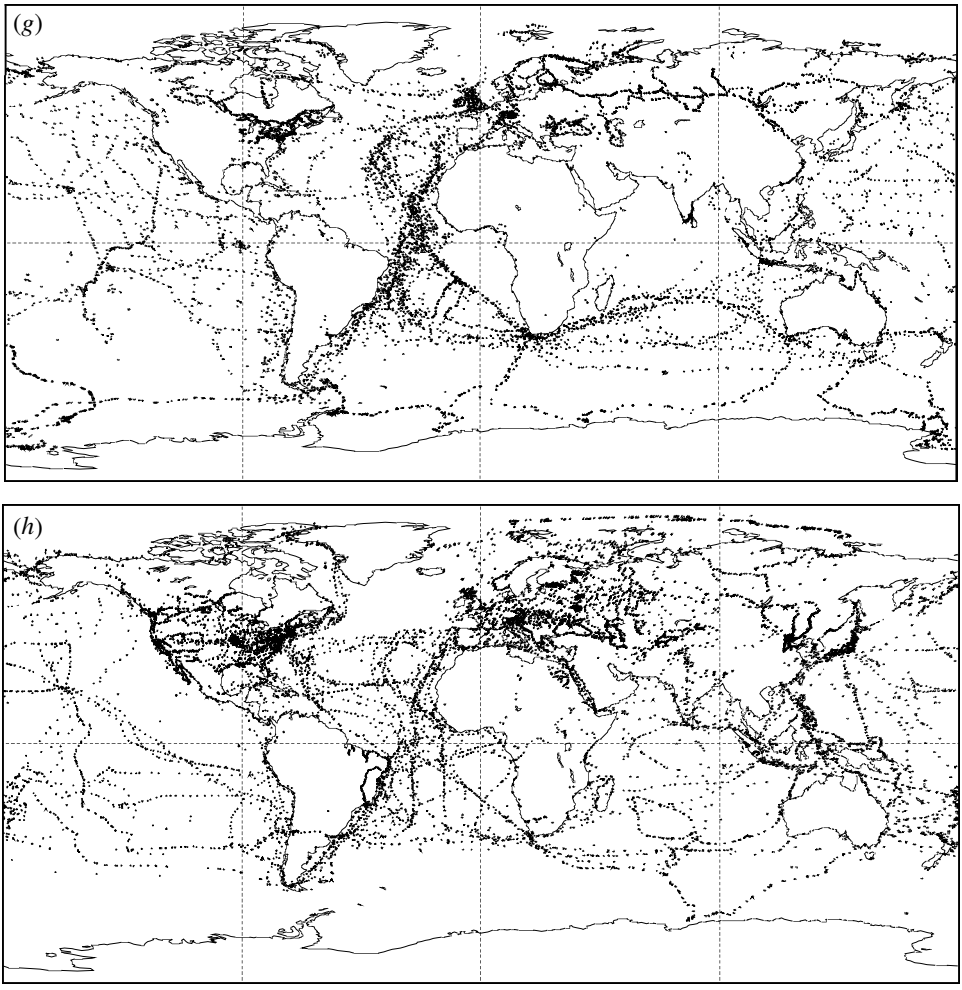

Figure 1. (*Cont.*) (*g*) All data 1800–1849. (*h*) All data 1850–1899.

shortcoming of the compilation is the fact that data were omitted by Sabine when forming his compilation from the original sources. Since the world compilations are by region rather than by voyage and are only for the zones 40–85° N, 0–40° N and 0–40° S, data from greater than 40° S have been omitted (notwithstanding the data originating in the Magnetic Survey of the South Polar Regions undertaken between 1840 and 1845, reported in Sabine (1868)). We used Sabine's dataset as the basis for creating a new dataset in which individual voyages are represented, and we used the original sources to reinstate missing data from the voyages. A more complete account will be forthcoming.

A major source of data for the early 19th century originates as two manuscript compilations held in the Archives Nationales and the Bibliotheque Nationale in Paris. The majority of the data are observations of declination made on board French Naval and Hydrographic Service ships. Since the data were not recorded in their original manuscript form, but had been transcribed onto a geographical grid, there

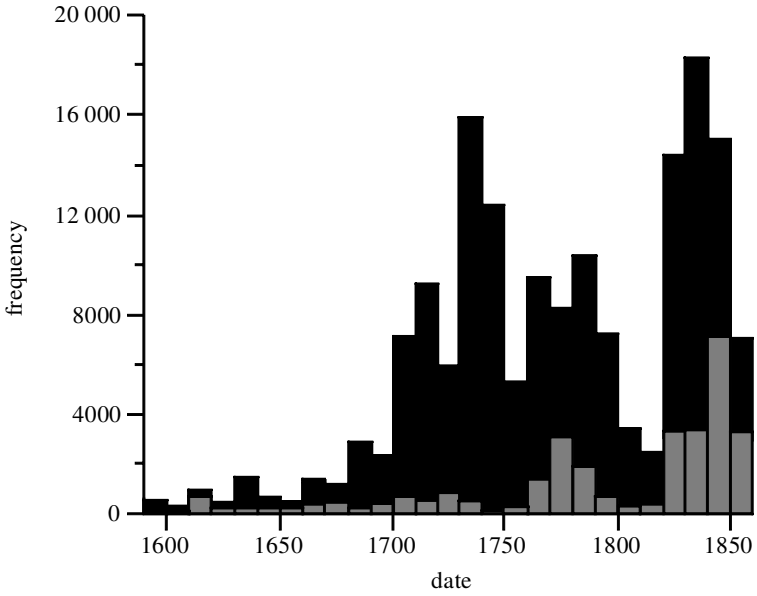

Figure 2. Histogram showing the temporal distribution of the new data collected (black), compared with the previous historical collection (grey) used in Bloxham & Jackson (1992).

is a quantization error in the observation positions that must be taken into account. Some observations are given on a $1^\circ \times 1^\circ$ grid, while others are on a $10^\circ \times 10^\circ$ grid. We have accounted for the positional imprecision in the same way as for pre-19th century data.

For the 20th century, we have used the same dataset as was used by Bloxham *et al.* (1989), which includes survey data, observatory data (which runs back to the mid-19th century), and data from the POGO and Magsat satellites. The temporal distribution of the final dataset up to 1860 is shown in figure 2, showing the large increase in data over that used in the production of the models *ufm1* and *ufm2* of Bloxham & Jackson (1992).

3. Method

(*a*) *Definitions and notation*

We begin by reviewing the concepts and notation necessary for our analysis. Much is standard and can be found, for example, in the texts of Langel (1987) or Backus *et al.* (1996). We adopt a spherical coordinate system (r, θ, ϕ), where r is radius, θ is colatitude and ϕ is longitude. We denote the Earth's radius and core radius by a and c, respectively, and take $a = 6371.2$ km and $c = 3485$ km. We assume the mantle is an insulator so that for $r > c$

$$\mathbf{B} = -\nabla V, \tag{3.1}$$

where V is the magnetic potential. Although other methodologies are possible, some of which are attractive for certain applications (e.g. Shure *et al.* 1982; Constable *et al.* 1993; O'Brien & Parker 1994), we use the spherical harmonic expansion of V as the underlying parametrization of the radial magnetic field at the CMB $B_r = -\partial_r V|_{r=c}$,

which, as we stated in § 1, is the quantity in which we are ultimately interested. At any time t the potential can be written

$$V = a \sum_{l=1}^{1} \sum_{m=0}^{l} \left|\frac{a}{r}\right|^{l+1} (g_l^m(t) \cos m\phi + h_l^m(t) \sin m\phi) P_l^m(\cos\theta), \tag{3.2}$$

where $\{g_l^m(t); h_l^m(t)\}$ are the so-called Gauss coefficients, and $P_l^m(\cos\theta)$ are Schmidt quasi-normalized associated Legendre functions. The choice of parametrization thus far has proceeded without loss of generality: the spherical harmonics represent a complete set. Computational considerations require that we truncate the spherical harmonic expansion at degree L; in this paper, as in previous work (e.g. Bloxham & Jackson 1992), we take $L = 14$. This choice is beyond the point in the spherical harmonic power spectrum, measured at the Earth's surface, at which the core field is overwhelmed by the crustal magnetic field. We also require a parametric representation for the time-varying Gauss coefficients. Here, we make a choice that restricts the class of functions that can be represented. With one eye on the regularization that we will introduce in equation (3.7) below, we choose the B-splines (e.g. Lancaster & Salkauskas 1986) of order 4 as basis functions for the expansion of the g_l^m and h_l^m, exactly as in Bloxham & Jackson (1992), wherein a rationalization can be found,

$$g_l^m(t) = \sum_n {}^n g_l^m B_n(t), \tag{3.3}$$

with a similar expansion for the $h_l^m(t)$. We choose to use T B-splines for the time period $[t_\mathrm{s}, t_\mathrm{e}]$, erected on a set of $T + 4$ regularly spaced knots. The $\{{}^n g_l^m; {}^n h_l^m\}$, when suitably ordered (we chose to let n increase the most slowly, then l, then m), constitute what we will call the model vector $\mathbf{m}$, with dimension $P = TL(L + 2)$ and components m_j.

Each component of the field constitutes a datum γ_j with an associated forward function $f_j(\mathbf{m})$ that makes a prediction from the model. Because many of our data are nonlinear functionals of the model $\mathbf{m}$, our methodology for finding a model from the observations proceeds by an iterative method, whereby we calculate small increments $\delta\mathbf{m}$ to the model at the ith iterate. We will frequently require the sensitivity of the jth synthesized datum with respect to a change in the kth model parameter m_k; this is simply the derivative $\partial f_j(\mathbf{m})/\partial m_k$. We store these derivatives in the matrix A with $A_{jk} = \partial f_j(\mathbf{m})/\partial m_k$.

We group the data into a data vector $\boldsymbol{\gamma}$ of length N, with an associated error vector $\mathbf{e}$. It is necessary to assign the e_i for each datum, depending on the methods of acquisition and the likely sources of error; this is given considerable attention in § 4. Of particular importance is the correlation in the errors, described by the covariance matrix C_e with elements

$$[C_\mathrm{e}]_{ij} = E\{e_i e_j\}, \tag{3.4}$$

where E signifies expectation. For most sources, the errors are independent, so that C_e is diagonal; this is the situation that is most commonly assumed in geophysics. However, one source of error, namely that due to imprecise navigation, is prone to lead to correlated errors because navigational errors made on a particular day compound errors made on previous days. We have developed a stochastic formalism

to account for this in our attribution of error; it leads to a non-diagonal C_e, although the correlation only affects data between landfalls within each individual voyage; this is described in § 4 *a*.

(*b*) *The continuous time–space model*

Before describing our solution methodology, it is pertinent to discuss the issue of uniqueness of our inverse problem: given perfect continuous knowledge of a property of the magnetic field on the surface of the Earth, we must consider if this is sufficient to reconstruct the magnetic field everywhere. We fix attention on one particular epoch. It is well known that perfect knowledge of the vertical component of the magnetic field is sufficient to determine the magnetic field everywhere, as a result of the uniqueness theorem for Laplace's equation with Neumann boundary conditions (e.g. Kellogg 1929). Indeed, knowledge of the horizontal field will suffice equally well. A different issue relates to whether field directions are sufficient to determine the magnetic field everywhere. This is particularly relevant, since our dataset is heavily biased towards declination measurements in the earliest years, and there are no absolute intensities prior to Gauss's invention of a method for determining intensity in 1832. The question has been recently addressed by Hulot *et al.* (1997), who found that perfect knowledge of the field direction on a surface is sufficient to determine the magnetic field everywhere, provided that the magnetic field has no more than two dip poles. An ambiguity that remains is that before 1832 our model can give only the field morphology, and requires a scale factor to fix its intensity. We adopt the dipole decay with time suggested by the extrapolation of Barraclough (1974); this fixes $g_1^0(t)$ to decay at a rate of 15 $\mathrm{nT\,yr^{-1}}$ for $t < 1840$ (essentially the rate of Barraclough (1974)), with the value in 1840 being determined by the intensities in the dataset.

A second and possibly more pertinent issue is that of non-uniqueness resulting from the finiteness of the dataset being analysed. It is well known (Backus & Gilbert 1970) that there are many models of a continuous function, such as B_r, that are capable of fitting a finite dataset. Therefore, we stress that in the methodology that follows we choose a model that satisfies our prejudices of being smooth in time and space, while still fitting the data. We do not claim any sort of uniqueness for the images that we present. Nevertheless, it is useful at this stage to mention the way that observations sample the core field. Figure 3 shows a cross-section of the (two-dimensional) kernel that dictates the sensitivity of a declination observation to departures of the core field from an axial dipole state. The point to note from this is that a declination observation is sensitive to perturbations in the core field over a large angular distance, and we can tolerate some gaps in our geographical data distribution without losing sensitivity to the core field entirely (see, for example, Johnson & Constable (1997) for more details).

Our methodology for regularizing the time–space model is the same as that presented in Bloxham & Jackson (1992). We find the smoothest model for a given fit to the data by seeking as our estimate the model vector $\mathbf{m}$ that minimizes the misfit to the data and two model norms, one measuring roughness in the spatial domain and the other roughness in the temporal domain. For the spatial norm, we seek the solution with minimum ohmic dissipation based on the ohmic heating bound of Gubbins

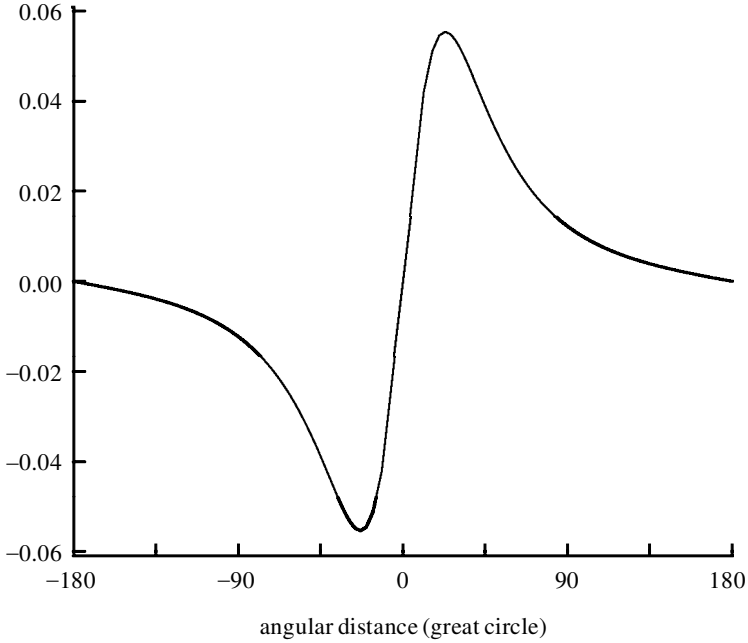

Figure 3. The kernel describing the sensitivity of a declination observation to the radial core field, when linearized about the field of an axial dipole. The ordinate shows the distance from the observation site, measured along the great circle passing through the site in an east–west direction.

(1975); we minimize the integral

$$\Psi = \frac{1}{(t_\mathrm{e} - t_\mathrm{s})} \int_{t_\mathrm{s}}^{t_\mathrm{e}} \mathcal{F}(B_r)\,\mathrm{d}t = \mathbf{m}^\mathrm{T} S^{-1} \mathbf{m}, \tag{3.5}$$

where $\mathcal{F}(B_r)$ is the quadratic norm associated with the minimum ohmic heating of a field parametrized by $\{g_l^m; h_l^m\}$:

$$\mathcal{F}(B_r) = 4\pi \sum_{l=1}^{1} \frac{(l+1)(2l+1)(2l+3)}{l} \left|\frac{a}{c}\right|^{2l+4} \sum_{m=0}^{l} [(g_l^m)^2 + (h_l^m)^2]. \tag{3.6}$$

For the temporal model norm, we use

$$\Phi = \frac{1}{(t_\mathrm{e} - t_\mathrm{s})} \int_{t_\mathrm{s}}^{t_\mathrm{e}} \oint_\mathrm{CMB} (\partial_t^2 B_r)^2\,\mathrm{d}\Omega \mathrm{d}t = \mathbf{m}^\mathrm{T} T^{-1} \mathbf{m}, \tag{3.7}$$

where $[t_\mathrm{s}, t_\mathrm{e}]$ is the time-interval over which we solve for the field.

We have assumed that the errors contaminating our dataset are Gaussian, with known covariance matrix C_e; this assumption leads, naturally, to the use of least-squares in the estimate of the model $\mathbf{m}$. We caution the reader that there is accumulating evidence that the distribution of errors that contaminate geomagnetic data may be far from Gaussian; therefore, we make our estimate robust to the presence of outliers by the use of a clamping scheme, whereby large residuals are gradually rejected from the estimation process until only data whose residuals are less

than three standard deviations away from the current prediction for that datum are retained (as used in Bloxham *et al.* (1989)). Our model estimate $\mathbf{m}$ is that which minimizes the objective function

$$\Theta(\mathbf{m}) = [\boldsymbol{\gamma} - \mathbf{f}\,(\mathbf{m})]^{\mathrm{T}} C_{\mathrm{e}}^{-1} [\boldsymbol{\gamma} - \mathbf{f}\,(\mathbf{m})] + \mathbf{m}^{\mathrm{T}} C_{\mathrm{m}}^{-1} \mathbf{m}, \tag{3.8}$$

where C_{e} is the data error covariance matrix (specified in § 4), and $C_{\mathrm{m}}^{-1} = \lambda_{\mathrm{S}} S^{-1} + \lambda_{\mathrm{T}} T^{-1}$, with damping parameters λ_{S} and λ_{T}. The solution is sought iteratively using the scheme

$$\mathbf{m}_{i+1} = \mathbf{m}_i + (A^{\mathrm{T}} C_{\mathrm{e}}^{-1} A + C_{\mathrm{m}}^{-1})^{-1} [A^{\mathrm{T}} C_{\mathrm{e}}^{-1} (\boldsymbol{\gamma} - \mathbf{f}\,(\mathbf{m}_i)) - C_{\mathrm{m}}^{-1} \mathbf{m}_i]. \tag{3.9}$$

A consequence of these regularization conditions is that the expansions (3.2) and (3.3) converge so that, with appropriately large values for the truncation points L and T, the solution is insensitive to truncation. We have used $L = 14$ and $T = 163$ in our calculations, with knots every 2.5 years. In addition to the regularization provided by (3.5) and (3.7), we have found it advantageous to impose boundary conditions at t_{s} and t_{e}. The so-called natural boundary condition of vanishing second (time) derivative has been applied at t_{s}, so that $\ddot{\mathbf{B}}(t_{\mathrm{s}}) = 0$. The equivalent condition at t_{e} has not been applied; doing so creates a large misfit, primarily to the observatory data. In effect, the observatory data supply a boundary condition on $\mathbf{B}$ themselves (recall that we use first differences of observatory data), and the pre-1990 secular variation is strongly at variance with the $\ddot{\mathbf{B}}(t_{\mathrm{e}}) = 0$ assumption. We have therefore refrained from applying this boundary condition at t_{e}.

Since the error covariance matrix is sparse (it is block diagonal), we can compute the contribution to the total misfit M from any subset of the data of size N_s with errors $\mathbf{e}^{(s)}$ as

$$M_s = \sqrt{\frac{1}{N_s} (\mathbf{e}^{(s)})^{\mathrm{T}} [C_{\mathrm{e}}^{(s)}]^{-1} (\mathbf{e}^{(s)})}, \tag{3.10}$$

where $C_{\mathrm{e}}^{(s)}$ is the $N_s \times N_s$ submatrix of C_{e} associated with the data sample; necessarily, it must encompass all correlations. The total misfit clearly satisfies

$$M^2 = \frac{1}{N} \sum_s N_s M_s^2, \tag{3.11}$$

a weighted sum over all the subset misfits.

(*c*) *Data correction*

After archival data conversion to machine-readable form and correction of obvious typographical errors, all of our new pre-19th century data have been manually examined by plotting in a purpose-built 'voyage editor', an interactive graphical user interface built around the GMT plotting package (Wessel & Smith 1991) designed for us by Dr N. Barber. The identification of geographical locations mentioned in the log is invaluable in enabling a reconstruction of the actual course plots; the voyage editor provided instant visual feedback on land sightings and allowed positional adjustments of three kinds: single-point editing; translation of voyage legs; and stretching of voyage legs.

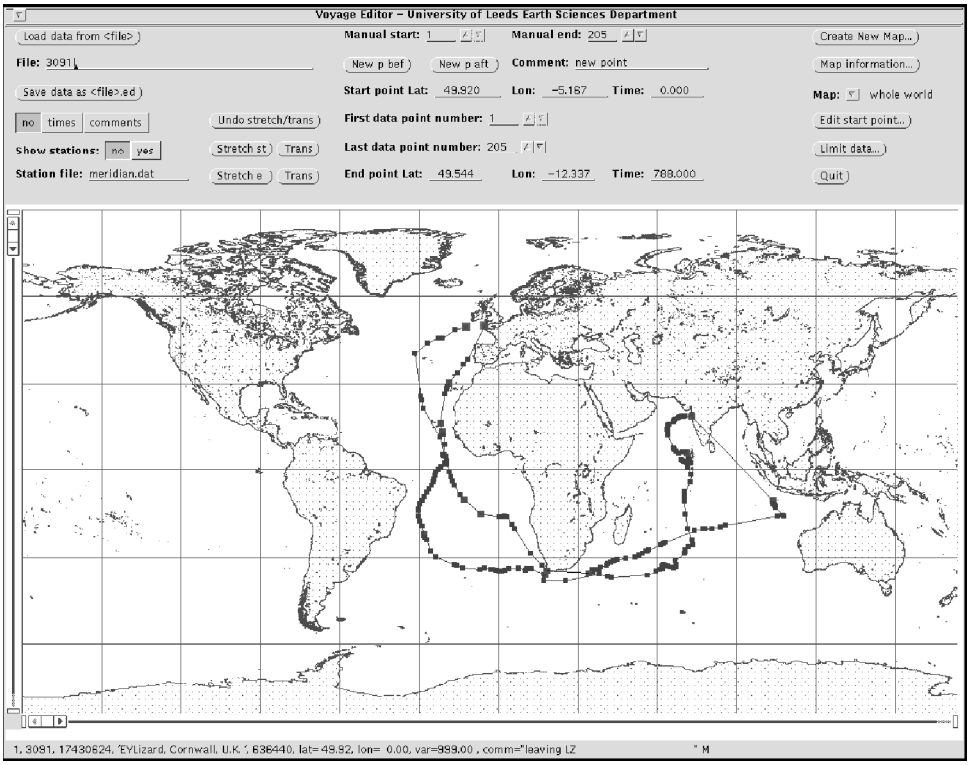

Figure 4. Example of a voyage loaded into the 'voyage editor'.

Translation and stretching operations acted on multiple points, which we refer to as a 'leg' of a voyage; a start and end point were chosen to define the leg boundaries, and either could be lifted and transported elsewhere. This displacement could then be effected uniformly for each point, shifting the whole leg to the right or left while maintaining latitude. Alternatively, in the case of stretching, one end of a leg remained fixed, while the other points were longitudinally stretched to meet the translated other end. For example, when a ship travelling eastward from Tristan da Cunha finally spotted the Cape of Good Hope, the westerly winds and currents carrying the ship along may have been underestimated or overestimated to such an extent that the reckoning put the vessel either far ahead or far before the longitude of Capetown. The last land previously sighted (the islands of Tristan) could then be fixed as a starting point, and the entry containing the comment 'seen the Cape of Good Hope', or 'at anchor in Simons Bay' moved to the appropriate location. The stretch was distributed equally over time, and all longitudes recalculated to proportionally reflect the change through linear transformation. This assumption, namely that the dead-reckoned error in longitude had been accumulated at an equal rate, is the same assumption as made by Bloxham (1986). An example of a voyage in the data editor is shown in figure 4.

4. Error estimates

A careful analysis of the sources of error in our data is essential if the data are to be used optimally. There are three major sources of error: navigational errors,

observational errors and crustal magnetic fields; we estimate the covariance matrices respectively as $C_{\rm n}$, $C_{\rm o}$ and $C_{\rm c}$. The final errors associated with the observations are given by the covariance matrix $C_{\rm e}$, which is the sum of all the contributing covariance matrices:

$$C_{\rm e} = C_{\rm n} + C_{\rm o} + C_{\rm c}. \tag{4.1}$$

(*a*) *Navigational errors*

Beginning with navigational errors, we turn to the issue of estimating the errors in the data that originate from the fact that prior to the use of the marine chronometer at sea, longitudes were estimated by the process of dead reckoning. As a result, navigational errors are often serially correlated. The estimate of longitude ϕ on day t_i is obtained by adding an estimated *increment*, σ_i, to the longitudinal estimate from the previous day:

$$\phi(t_i) = \phi(t_{i-1}) + \sigma_i. \tag{4.2}$$

The error in the daily increment, denoted $\delta\sigma_i$, is responsible for the serial correlation. We assume that $\delta\sigma_i$ is a zero-mean Gaussian random variable. Therefore

$$E\{\delta\sigma_i\} = 0, \tag{4.3}$$

and the variance of the daily error is

$$E\{\delta\sigma_i\delta\sigma_j\} = \varepsilon^2\delta_{ij}. \tag{4.4}$$

Here, ε is measured in degrees.

The covariance matrix of the longitudinal errors is

$$[C_\phi]_{ij} = E\{\delta\phi(t_i)\delta\phi(t_j)\}. \tag{4.5}$$

Usually, in addition to the dead-reckoned longitudes we know the initial and final port longitude, and often have intermediate land sightings. In the case of land sightings having been made, the voyage is divided into legs, with each leg treated separately. At each positional fix, we can compare the known position with the dead-reckoned position, giving an estimate of the accumulated error in a leg. These errors can be used to check the consistency of our random-walk model of navigation errors.

Two situations arise in the logs. The first is a leg that starts from a known position and a number of observations are made before the leg finishes with no geographical fix; only dead-reckoned longitudes are given. We will refer to this first class as a classical *random walk* (for the continuous time case) or *Brownian motion*. The second class comprises legs with a known initial position and a known final position, with intervening observations. This second case is the so-called *Brownian bridge* (e.g. Bhattacharya & Waymire 1990, ch. 1).

We begin by considering the case of the classical random walk. As is well known, the walk is a stochastic process resulting from the accumulation of *independent increments*. This is applicable because it is the error in the dead reckoning that contributes cumulatively to the positional uncertainty. For the random walk beginning at time t_0 and ending at time t_{n+1}, we have the result

$$[C_{\rm rw}]_{ij} = E\{\delta\phi(t_i)\delta\phi(t_j)\} = \varepsilon^2 \min(t_i, t_j), \tag{4.6}$$

simply because

$$\delta\phi(t_i) = \delta\phi(t_j) + \sum_{k=j+1}^{i} \delta\sigma_k \quad (\text{for } t_i > t_j), \tag{4.7}$$

and all the errors in the increments $\delta\sigma_k$ are uncorrelated. We note that the variance of the process increases as

$$\mathrm{Var}(\delta\phi(t)) = E\{\delta\phi(t)^2\} = \varepsilon^2 t, \tag{4.8}$$

which encapsulates the well-known result that the root-mean square deviation of a random walk increases as $\varepsilon\sqrt{t}$. Clearly

$$[C_{\mathrm{rw}}]_{0j} = [C_{\mathrm{rw}}]_{j0} = 0, \quad \forall j, \tag{4.9}$$

since the position is known perfectly at time t_0; without any other sources of error, this covariance matrix is improper, in as far as it is non-positive definite. However, the full covariance matrix is the sum of the covariance matrices of all the contributing sources of error, which removes this problem.

In the second case of the Brownian bridge (a known arrival point exists), the data from each leg are corrected in the manner described in § 3 *c*, using the purpose-built 'voyage editor'. The error at time $T = t_{n+1}$ is annihilated by bringing the longitude $\phi(T)$ into agreement with the known longitude; all intervening dead-reckoned longitudes are corrected by linear interpolation in time. This transformation generates the so-called Brownian bridge. More specifically, following Bhattacharya & Waymire (1990), let $\delta_\phi(t)$ be the error in longitude resulting from a standard Brownian motion starting at $\phi = 0$ at $t = 0$. We define

$$\widetilde{\delta_\phi}(t) = \delta_\phi(t) - \left|\frac{t}{T}\right| \delta_\phi(T), \tag{4.10}$$

giving $\widetilde{\delta_\phi}(0) = \widetilde{\delta_\phi}(T) = 0$, as required. We have

$$E\{\widetilde{\delta_\phi}(t)\} = 0, \tag{4.11}$$

and (for $t_i \leqslant t_j$ to make the issue definite) it follows from (4.10) that

$$\begin{aligned} E\{\widetilde{\delta_\phi}(t_i)\widetilde{\delta_\phi}(t_j)\} &= E\{\delta_\phi(t_i)\delta_\phi(t_j)\} - \left|\frac{t_j}{T}\right| E\{\delta_\phi(t_i)\delta_\phi(T)\} \\ &\quad - \left|\frac{t_i}{T}\right| E\{\delta_\phi(t_j)\delta_\phi(T)\} + \left|\frac{t_i t_j}{T^2}\right| E\{\delta_\phi(T)\delta_\phi(T)\} \\ &= \varepsilon^2 \left| t_i - \left|\frac{t_j}{T}\right| t_i - \left|\frac{t_i}{T}\right| t_j + \left|\frac{t_i t_j}{T^2}\right| T \right| \\ &= \varepsilon^2 t_i (1 - t_j/T), \quad \text{for } t_i \leqslant t_j, \end{aligned} \tag{4.12}$$

and, thus, we define the Brownian bridge covariance matrix C_{bb} as

$$[C_{\mathrm{bb}}]_{ij} = E\{\widetilde{\delta_\phi}(t_i)\widetilde{\delta_\phi}(t_j)\} = \varepsilon^2 t_i (1 - t_j/T), \quad \text{for } t_i \leqslant t_j. \tag{4.13}$$

Equation (4.13) shows that

$$[C_{\mathrm{bb}}]_{0j} = [C_{\mathrm{bb}}]_{j0} = [C_{\mathrm{bb}}]_{j(n+1)} = [C_{\mathrm{bb}}]_{(n+1)j} = 0, \quad \forall j, \tag{4.14}$$

since the position is known perfectly at both the start and end points. It should also be noted that the variance of the error varies like $\varepsilon^2 T\tilde{t}(1-\tilde{t})$, where $\tilde{t} = t/T$. The variance is maximum at the mid-point of the journey, just as it should logically be; note, however, that the amplitude of the maximum error has been reduced, compared with the ordinary random walk (which scales as $\varepsilon^2 T\tilde{t}$), by a factor of two when the standard deviations are compared; the knowledge of the correct end point is valuable.

To convert these positional errors into errors in the observation, we multiply by the gradient of the field in the longitudinal direction, $\partial D/\partial\phi$; then the contribution of the navigational errors to the total covariance matrix reads

$$[C_{\mathrm{n}}]_{ij} = [C_\phi]_{ij} \frac{\partial D(\phi_i)}{\partial \phi} \frac{\partial D(\phi_j)}{\partial \phi}, \tag{4.15}$$

where C_ϕ is either of C_{bb} or C_{rw} depending on whether the leg ended at a landfall or not. We compute the gradients using the estimate of the magnetic field model at the current iteration. In general, $\partial D/\partial\phi$ is somewhat less than a degree per degree of longitude, although, of course, errors in longitude near to the magnetic poles can translate into considerable errors in D because of the huge gradients there.

This simple method of converting positional errors into the equivalent error in the datum is a pragmatic approach to what would otherwise be a very complicated problem. Of the other approaches to dealing with errors in the *independent* rather than the *dependent* variable, the approach of treating the positions as parameters to be determined along with the geomagnetic field model has 'the makings of a very poorly determined inverse problem', according to Bloxham (1985); a second approach, designed to deal with errors in the design matrix (which is exactly the case here), is termed the 'total least-squares problem' and remains a subject of research (see, for example, Van Huffel & Vandewalle 1991). Suffice it to say that we do not believe that the computational obstacles that would have to be overcome would lead to radically different solutions to those determined here.

We can use the voyages themselves to test the validity of the random-walk model. At the first landfall after the start of a leg we estimate the accumulated error in longitude; according to the model, each of these errors is a sample from $\delta_\phi(t)$ after time t. A plot of $\delta_\phi(t)$ as determined from the voyages is given in figure 5*a*; the data contributing to this plot are those corrections that were manually performed using the voyage editor, based on objective information regarding the final position. We have omitted those voyages whose legs had no corrections, a necessity forced on us by the fact that some logs have been left uncorrected merely because we lacked any information on their recorded landfall. If included, these logs would create an unduly large peak at zero error, completely biasing the analysis. It is, therefore, an unfortunate state of affairs that we cannot separate logs with perfect navigation (which must occur on a small number of occasions) from those for which navigational correction was impossible.

We would like to check that some of the underlying features of the random-walk model are exhibited by the data, bearing in mind that exact correspondence is unlikely. The random-walk model requires that $E\{\delta_\phi(t)\} = 0$, which is reasonably well satisfied by the errors; this is important as it suggests that systematic errors are probably not present in the data. We recall that there was at least rudimentary knowledge of ocean currents that might well have been accounted for by the navigators. We can also check that the width of the distribution increases with time, as

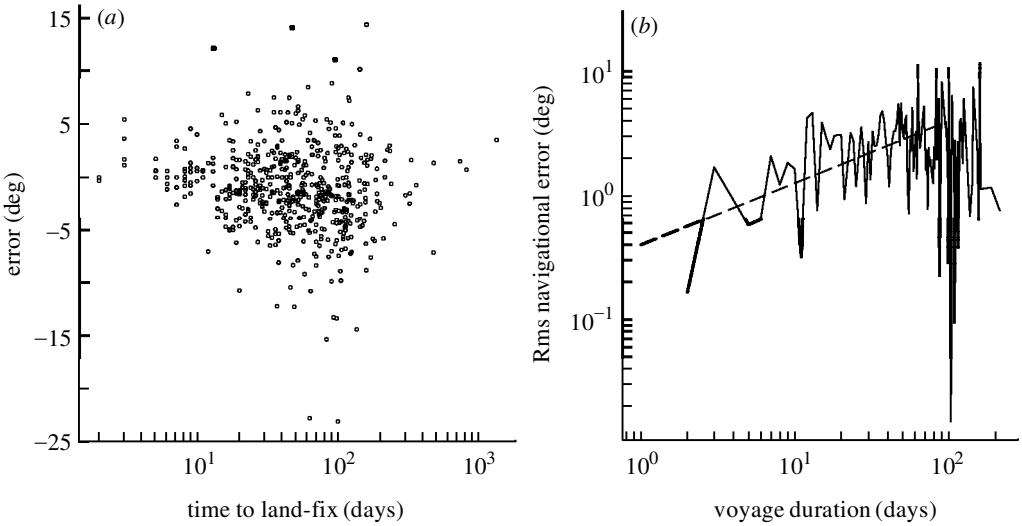

Figure 5. (*a*) Scatterplot showing the individual errors in longitude accumulated as a function of time for the voyages we corrected using land sighting information. (*b*) The data in (*a*) as a function of time for all days with two or more samples. The dotted line is the theoretical prediction for a $\sqrt{t}$ increase with time t if $\varepsilon = 0.4^\circ$.

predicted by the random-walk model. To illustrate this, we analyse the data as a function of the length of the voyage, using days on which two or more samples are present. Figure 5*b* shows the RMS deviation as a function of the length of the voyage in days. Reassuringly, we find that the variance increases as a function of time, and there is a reasonable agreement with the square-root-of-time model predicted by the random-walk model, although it is unfortunate that there are few long voyages (or legs), and, thus, there is considerable scatter for $t > 100$ days. The statistics are biased upwards by the lack of voyages with zero or small corrections, and we must look for another indicator of the rate at which longitudinal error is accumulated. We turn to the dead-reckoned latitudes.

It is not uncommon for both dead-reckoned latitudes and astronomically determined latitudes to be recorded in a log, and we have a considerable number of these duplicate positional estimates in our database. We have used Dutch logs containing both dead-reckoned and astronomical latitudes to determine the latitudinal errors from the difference in the two readings. The difference between the errors in latitude and the corresponding errors in longitude is in the frequency of the accurate determinations of position (i.e. astronomical determinations of latitude and landfalls/sightings in the case of longitudes). It was common for astronomical latitudes to be found daily, weather permitting, and although there are instances of storms or overcast weather obscuring the sky for up to a week, it was rare for a ship to go for more than a few days without a latitudinal fix; either the sun or the stars could almost always be observed to enable a fix to be made. The errors in latitude represent, overwhelmingly, errors accumulated over one day, with a few representing errors over a handful of days. (From the records, it is difficult for us to determine the time since the navigator made his last astronomical fix because we only record days on which magnetic data are recorded; to do otherwise could have increased

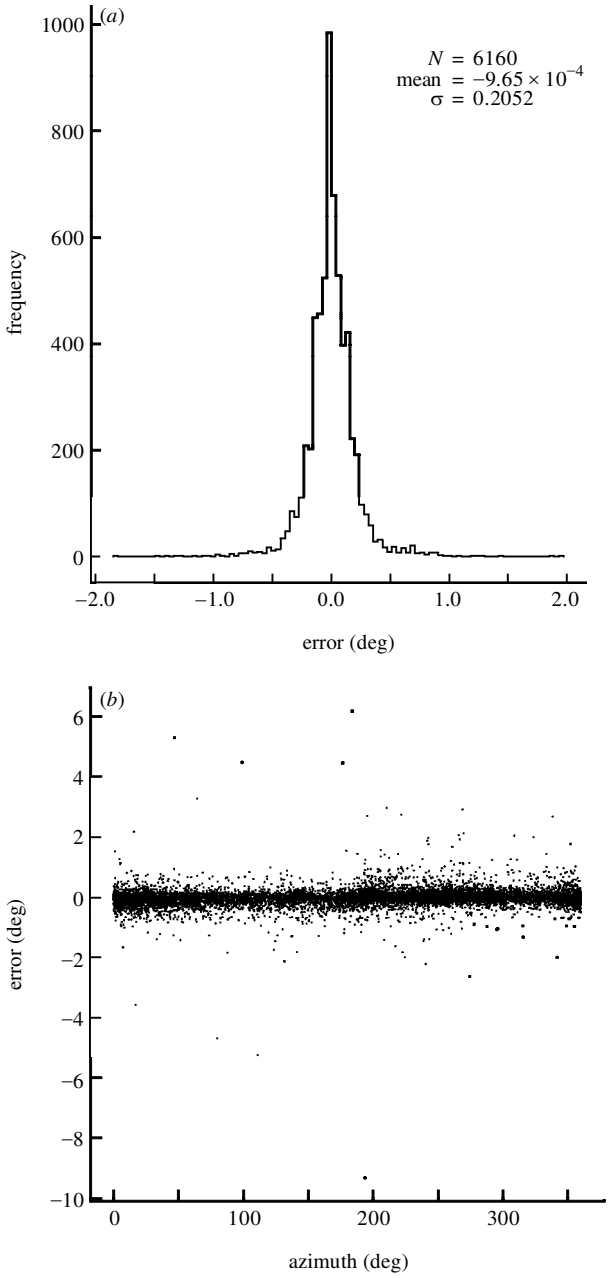

Figure 6. (*a*) Differences in dead-reckoned and astronomically determined latitudes. (*b*) Same as (*a*) but as a function of heading.

the data-entry task by up to an order of magnitude.) Figure 6*a* shows a histogram of the latitude errors from our dataset that have been taken within one day of a previous magnetic measurement. We can see that the errors in latitude are equally distributed about zero, and have a standard deviation of approximately 0.2°. Part of

this error distribution originates in inaccuracies in the dead-reckoned latitudes, and part originates in errors in the astronomical determinations. We estimate that the astronomical determinations are good to about 10° of arc, so it is possible that half of the variance observed comes from this source. This figure is extremely robust: when all available dead-reckoned latitudes are analysed (including those which may have been made after an interval of more than one day had passed), we find a standard deviation of 0.23° from a total of 11 670 observations.

We have examined the distribution of these errors as a function of the course on which the ship was sailing, but found no significant variation (figure 6*b*). If the main source of error was due to inaccuracy in the estimated distance that the ship had run, we would expect larger errors on north–south voyages than on east–west voyages. The null result suggests that errors originate from the combination of misestimates of distance run combined with drift. The fact that the errors are isotropic means that we can reasonably make the assumption that they are a good estimate of the daily error in dead-reckoned longitudes. The biased figure for the daily error, obtained from the analysis of the navigational corrections, was of the order of 0.4°. We have good grounds for rejecting this figure as too large, and instead consider the analysis of the latitudinal dead reckoning to be a more useful indicator. However, we are loathed to reduce it to a figure of 0.16°, which would be appropriate if we assign part of the observed errors to astronomical error. Therefore, we use the standard deviation obtained here to fix $\varepsilon = 0.2^{\circ}$. This figure is used in our assignment of C_{n} via (4.15).

(*b*) *Observational errors*

We have been able to quantify observational errors in pre-19th century data. Contemporary accounts suggest that defects seem to have plagued the instruments used by mariners of all nations sampled. In addition to technical limitations and a host of secondary ills, there were three major complaints: the needle being too weakly magnetized; the card not being properly oriented (skewed or warped); and the cap and pin being subject to friction and wear. Observational error constituted a second category, comprising atmospheric conditions (unclear sighting, refraction), parallax (both in sighting and reading off the dial), ship's movement, deviation, and human error in performing the task. As a rule, discrepancies between consecutive or simultaneous measurements of D with a magnitude perceived to be large enough to merit special mention in logs decreased through time, from a couple of degrees early in the 17th century to a few minutes late in the 18th. On several ships, careful analysis was made of performance differences, mostly between the official issue compass (by the Admiralty or EIC) and an alternative in the personal possession of the observer, which often proved superior.

Repeated observations on one day can be used to determine the observational error. For any day on which there are n repeated observations $\{\gamma_i\}$, we calculate the sample mean, $\hat{\mu}$, and sample variance, s^2, of the observations:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\gamma_i, \tag{4.16}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\gamma_i - \hat{\mu})^2 = \frac{1}{n-1}\sum_{i=1}^{n} r_i^2, \tag{4.17}$$

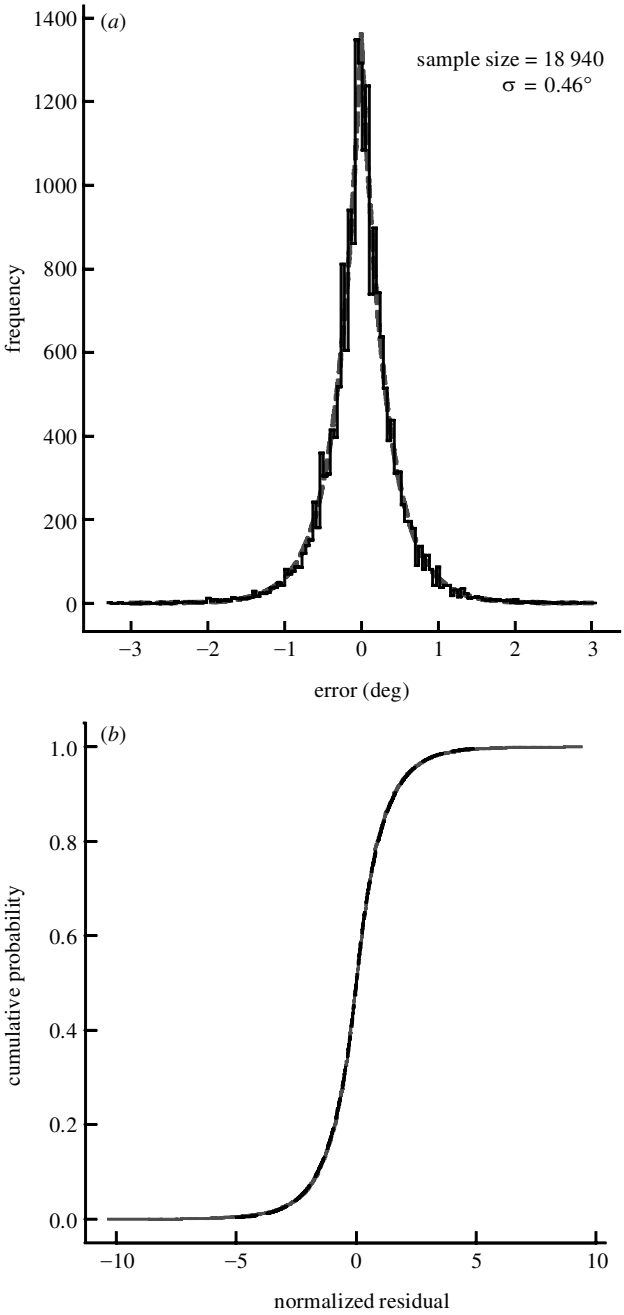

Figure 7. (*a*) Errors in repeated observations of declination made on the same day, plotted as a histogram. Also plotted (dashed) is the theoretical probability density function (PDF) for a Laplace distribution with parameter $\beta = 0.32^\circ$. (*b*) Same as (*a*) but the cumulative density function (CDF) is plotted. Also plotted is the theoretical CDF for a Laplace distribution with parameter $\beta = 0.32^\circ$; since the standard deviation of this distribution is $\sqrt{2}\beta = 0.45^\circ$, the sample standard deviation s (0.46°) and theoretical standard deviation agree almost exactly.

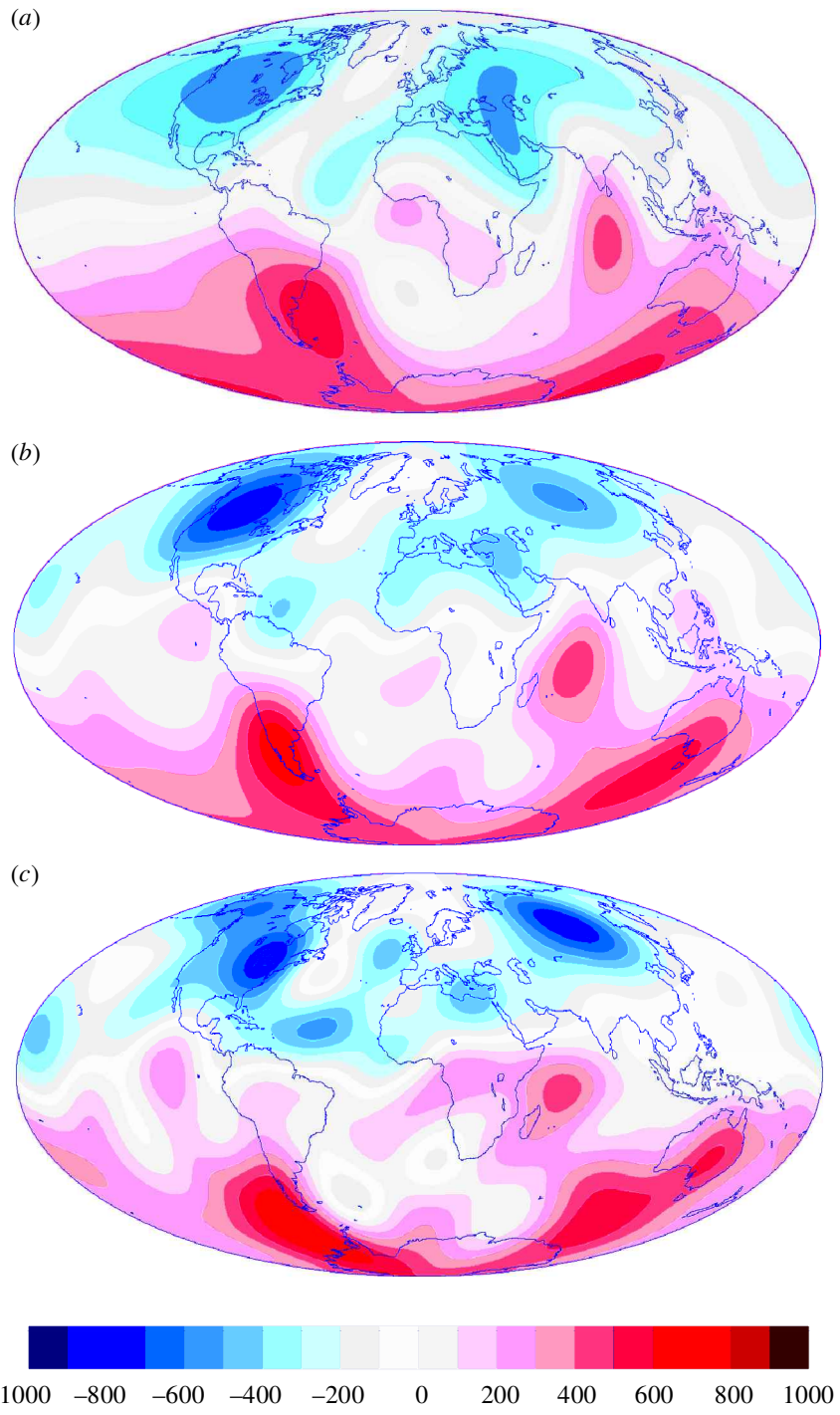

Figure 8. The radial field at the CMB on an Aitoff equal-area projection. The contour interval is 100 μT. Blue shades represent flux into the core, red shades flux out of the core. (*a*) 1590. (*b*) 1690. (*c*) 1790.

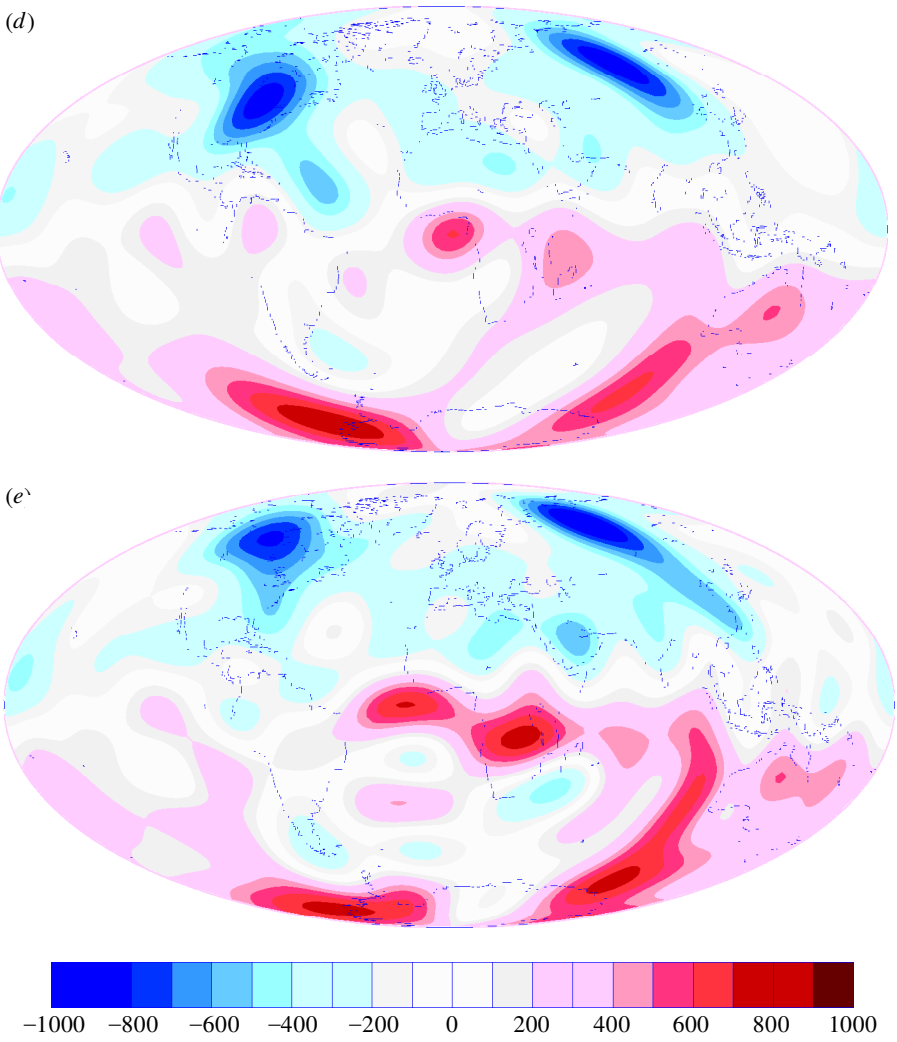

Figure 8. (*Cont.*) (*d*) 1890. (*e*) 1990.

where $\{r_i\}$ are the residuals from the calculated mean $\hat{\mu}$. We are interested in the error σ_μ on the true mean of the observations μ; this is well known to be given by

$$\sigma_\mu^2 = E\{(\mu - \hat{\mu})^2\} = s^2/n. \tag{4.18}$$

We use $\hat{\mu}$ as the single datum for each day, and assign it an error $s/\sqrt{n}$, but we determine a single value for s to apply for all the observations as follows. From the observations database, a total of 18 918 residuals r_i from their daily mean can be derived from days on which 2–4 measurements (or measurement sessions) were taken at different times during a single day. From these we find $s = 0.46°$.

Figure 7*a* shows a histogram of the observed errors. The errors are distinctly non-Gaussian; we have no explanation for this at present. Indeed, the data are very well represented by a Laplace or double-exponential distribution with probability

distribution function

$$p(x) = \frac{1}{2\beta} \mathrm{e}^{-jxj/\beta}. \tag{4.19}$$

When the cumulative distribution functions are compared (figure 7*b*), the agreement is spectacular. Despite this non-Gaussian behaviour, we have taken the value of $s = 0.46^\circ$ as representative of observational error before 1800; this value then forms the (diagonal) entry in the covariance matrix C_o.

(*c*) *Crustal magnetic field errors*

The presence of magnetized rocks at the Earth's surface constitutes, for our purposes, a source of noise in imaging the core field. Over the last ten years some progress has been made in characterizing this noise by assuming that the crustal field is a stationary isotropic random process (Jackson 1990, 1994, 1996; Backus 1988, 1996; Ma 1998). Such statistical models are useful for our purposes because it is almost certainly impossible to develop a deterministic model of the crustal field with sufficient fidelity that every magnetic anomaly is represented, down to sub-kilometre scales. Even if such a model existed, it would be difficult to use with our historical dataset, because of the possibility of navigational inaccuracy. Without precise positional certainty, the subtraction of a global crustal model could actually have the effect of increasing the noise in the data, as a result of subtracting the wrong anomaly; a statistical model is much more suitable for our purposes. In the spatial domain the statistical models describe a covariance function, giving the expected value of the correlation of the different magnetic field components (or equivalently the magnetic potential) at different positions on the Earth's surface. Such a covariance function can be used in the pre-whitening of the data, in a way analogous to the treatment of the navigational errors. The covariance function can, in principle, be discovered from the power spectrum of the magnetic field; this is equivalent to the idea that in a plane geometry a white power spectrum results from random noise that is purely uncorrelated, or has a delta function as its covariance function. Despite considerable progress in the theory, it remains difficult to determine the covariance function. By far the most useful observations must come from close to ground level, because of the superior resolving power of the data compared with satellite observations. Our geological prejudices dictate that the half-width of the correlation function must be rather small, of the order of a few tens of kilometres, possibly a few hundreds of kilometres at the very most. Calculations that take the correlation function into account in the analysis of satellite (Magsat) data (Rygaard-Hjalsted *et al*. 1997) found very small changes in the model estimated, mainly because the random errors in the data are significantly larger than the crustal errors (Jackson 1990), making the matrix diagonally dominant. This may not be the case for the new generation of satellites, such as Ørsted, and, therefore, the effect may be required to be re-analysed. A covariance matrix designed to account for the correlation caused by the crust has never been applied to surface data. To do so would be a massive computational undertaking. It is not clear to what extent we may be susceptible to aliasing the possibly correlated noise into long-wavelength core magnetic fields.

Amongst the observational constraints that appear indisputable is the fact that on average the vertical magnetic field has a larger magnitude than the horizontal

Table 1. *Statistics of the model*

number of data retained	365 694
number of data rejected	26 893
misfit	1.16
damping parameter λ_s (nT^{-2})	1×10^{-12}
damping parameter λ_T ($\mathrm{nT}^{-2}\ \mathrm{yr}^4$)	5×10^{-4}
spatial norm Ψ (nT^2)	35×10^{12}
temporal norm Φ ($\mathrm{nT}^2\ \mathrm{yr}^{-4}$)	6.8×10^4
RMS secular variation (at CMB) ($\mathrm{nT\ yr}^{-1}$)	1817

components of the field. The stochastic model predicts a value of $\sqrt{2}$ for the ratio of the RMS value of the Z component to the RMS values of H, X or Y, and we have adopted values for the amplitudes that exhibit this so-called anisotropy (Holme & Jackson 1997) of 200 nT (for $\sigma_{\mathrm{cr}}^{\mathrm{H}}$) and 300 nT (for $\sigma_{\mathrm{cr}}^{\mathrm{Z}}$); the theoretical value is not exactly reproduced, but the applicability of the isotropic model is not undisputed.

With these values of the crustal amplitudes, we generate the well-known standard deviations for the D and I measurements of

$$\sigma_{\mathrm{cr}}^{\mathrm{D}} = \sigma_{\mathrm{cr}}^{\mathrm{H}}/H, \qquad \sigma_{\mathrm{cr}}^{\mathrm{I}} = \sigma_{\mathrm{cr}}^{\mathrm{Z}}/F. \tag{4.20}$$

The values form the appropriate diagonal elements of C_{c}, when the required values of H and F are calculated at the desired location from the current iterate of $\mathbf{m}$.

We note here that the treatment of observatory data must take into account the presence of the crust, because otherwise the errors are liable to be correlated in time. We adopt the same approach as in Bloxham & Jackson (1992), namely that of using first differences of the data, which should be free of the effects of the crust if it is strictly constant in time.

5. Results

We present our results as a series of snapshots of the field at the core–mantle boundary through time, and give statistics of model *gufm1* in table 1.

Figure 8*a–e* shows the evolution of the field at 100 year intervals from 1590–1990. The results show good consistency with those produced previously from much smaller datasets, although there are some differences. The solutions exhibit the long-term features of the field that have become familiar since the work of Bloxham & Gubbins (1985); for example, the four large flux lobes at high latitudes over Canada and Siberia and their counterparts almost symmetrically placed to the south of the Equator, and the low intensity of flux at the North Pole, commonly believed to be indicative of the dynamical effect of the inner core. We appear to have imaged the Indian Ocean core spot that is visible in 1590 and which gradually drifts west over time to a final situation in Central Africa in 1990 particularly well. Additionally, the time dependency of the Canadian flux lobe should be noted: this is far from a static feature, splitting into two parts by the mid-19th century and exhibiting wave-like motion that is most easily viewed in the form of a movie made from a series of images.

In order to see the similarities and differences between this model and the previous models, *ufm1* and *ufm2*, in figure 9*a–c* we show three comparisons with the model

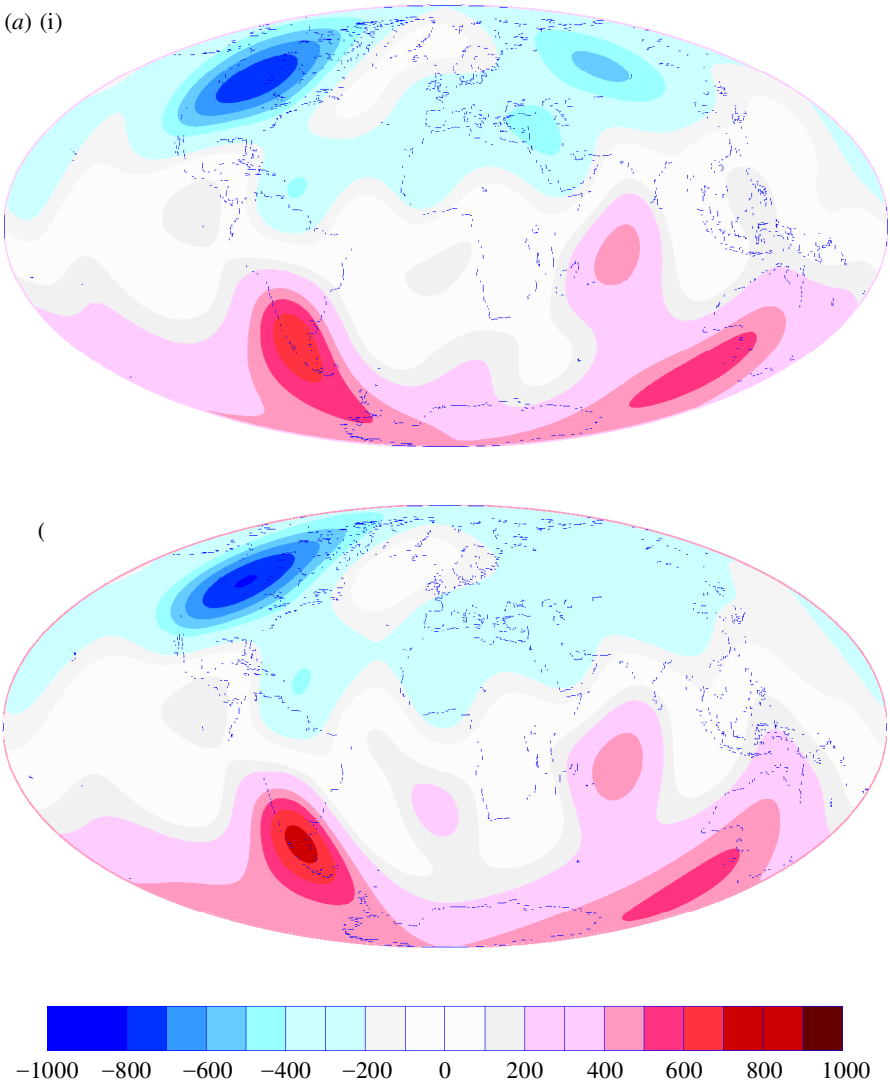

Figure 9. Comparisons of the radial field at the core–mantle boundary from the new model and model *ufm2* of Bloxham & Jackson (1992). (*a*) For 1690, *gufm1* (top) and *ufm2* (bottom).

ufm2 for the years 1690, 1765 and 1840. In 1690 (figure 9*a*), the flux lobe over Siberia, which had diminished amplitude in *ufm2*, is present. In 1765 (figure 9*b*), the reverse flux patch at the North Pole is slightly more prominent. Note that the splitting of the Canadian flux lobe is highly evident in figure 9*c*. This surprising similarity between the solutions probably results from the fact that the solutions we present are smoother than they should be, a problem that we discuss below.

Our model solves the optimization problem posed by (3.8) with the data covariance matrix C_e and the regularization matrix C_m prescribed by the preceding discussion. The solution is, therefore, 'optimal' in this sense. It is well known that the regularized solution can be interpreted in a Bayesian sense as being the mode of the

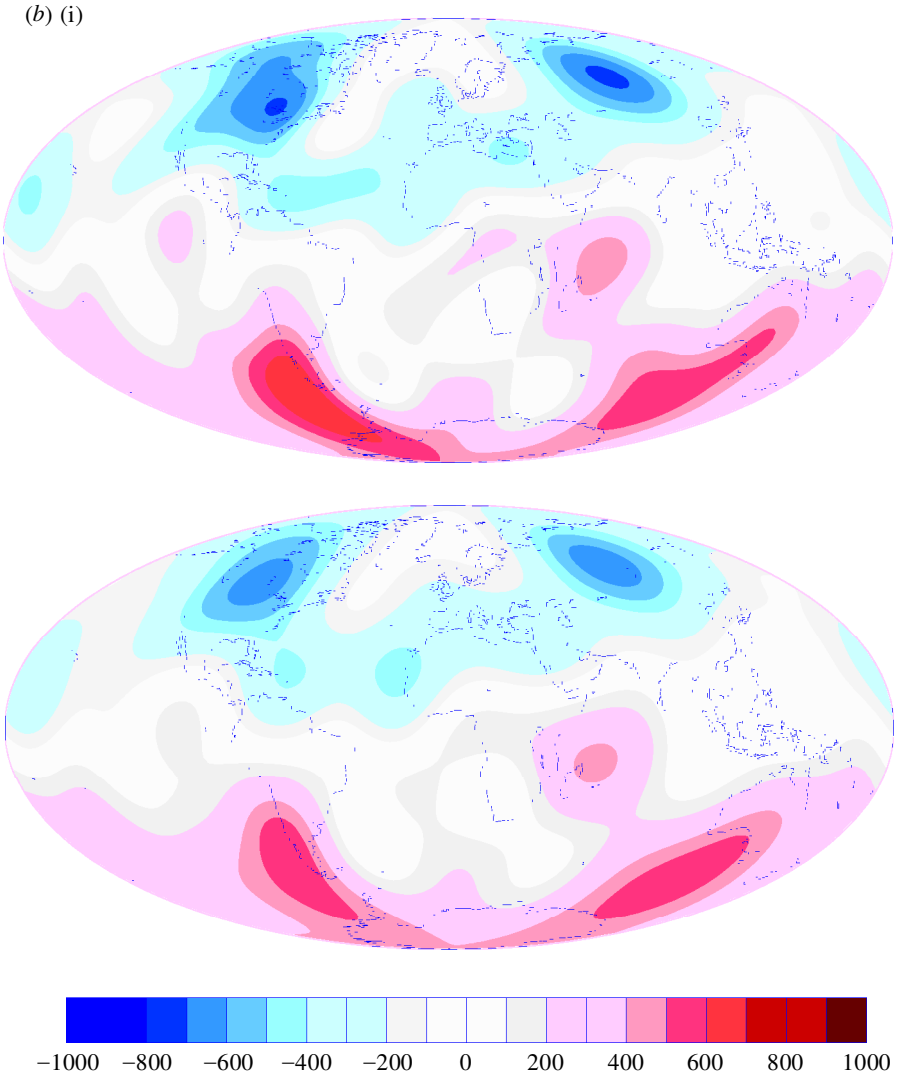

Figure 9. (*Cont.*) (*b*) For 1765, *gufm1* (top) and *ufm2* (bottom).

posterior probability density function; a discussion can be found in Bloxham *et al.* (1989). We caution the reader that this methodology for generating a solution has some features that we consider mildly undesirable. Figure 10*a* shows the misfit M_s of 5 year subsets of data as a function of time. The misfit increases slightly, on average, back in time, indicating that the solution fits the more recent data more closely. This effect is clearly linked to the number of data contributing to the solution as a function of time, shown in figure 10*b*. We also find that the spatial norm of the solution exhibits a similar increase with time, linked to the increase in the number of data (figure 10*c*). In a Bayesian sense, this is precisely what is to be expected: increased information content maps into more complex solutions, and when there is little information injected from the data, the solution is controlled by the prior information. This state of affairs is not tremendously satisfying, as discussed in Bloxham *et al.*

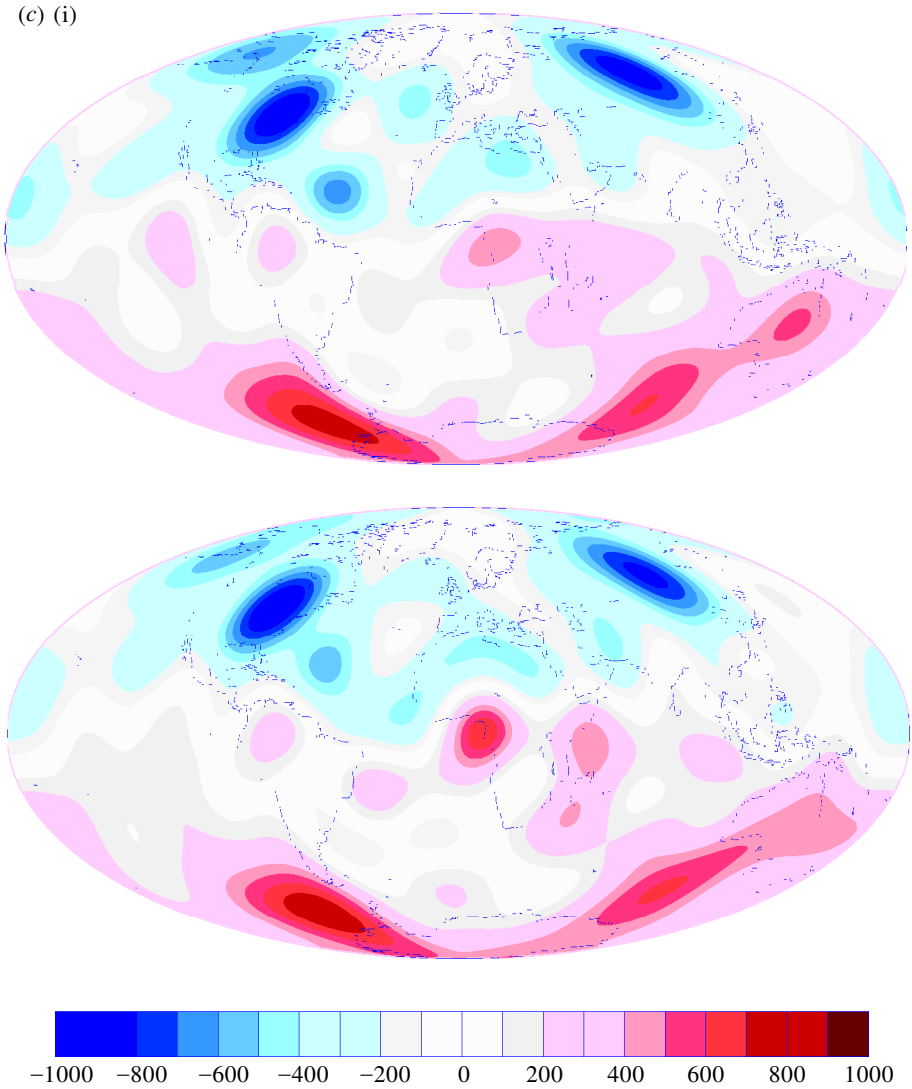

Figure 9. (*Cont.*) (*c*) For 1840, *gufm1* (top) and *ufm2* (bottom).

(1989); there is much to be said for retaining a similar number of degrees of freedom in the model as a function of time, although even this must be done at the expense of increased variance of the model estimate back in time. The alternative strategy, of actually winnowing the dataset to give poorer resolution in this century by using similar numbers of data to other times, does not look particularly attractive.

It would be possible to pose an optimization problem that would be capable of overcoming some of these 'undesirable' properties of the solution; the approach of fixing the spatial norm in time by using time-varying damping parameters comes to mind, and the approach developed by Love & Gubbins (1996) for the 'optimized dynamo problem' would certainly be a useful approach. We plan to look at this approach, and others, in the future.

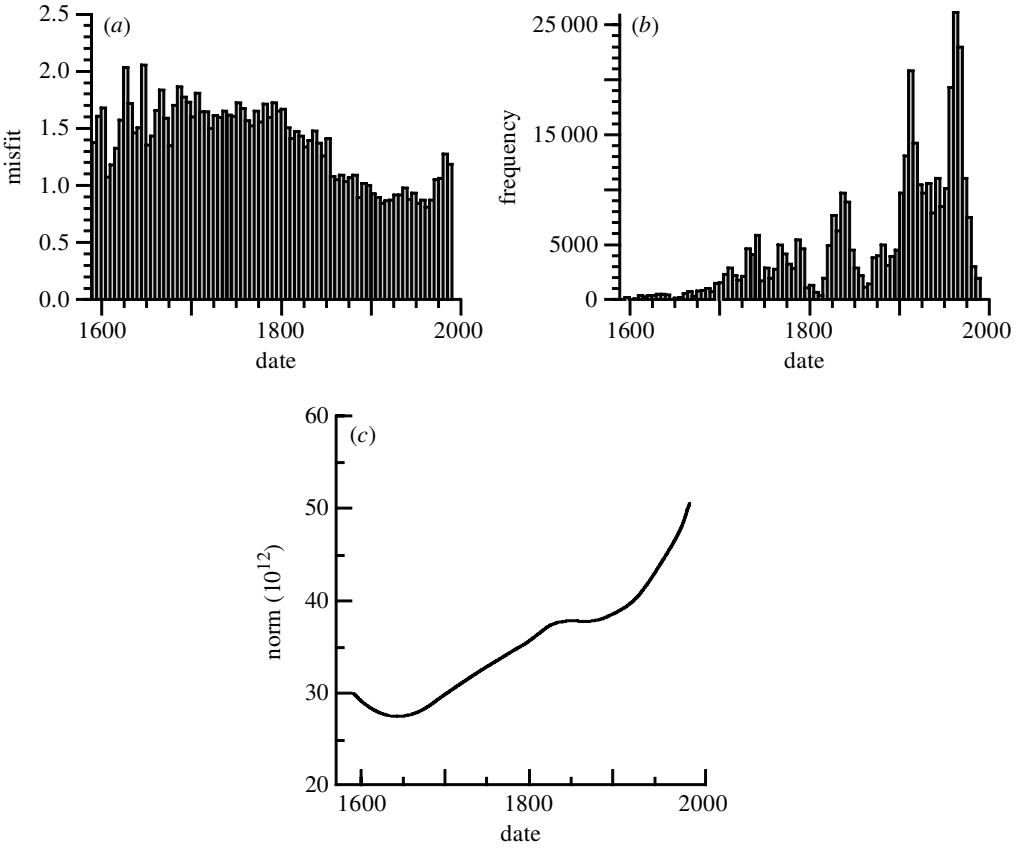

Figure 10. (*a*) Root-mean-square weighted misfit M_s between the model and the data for 5 year subsets of data. (*b*) The overall number of data in the model in 5 year bins. (*c*) Spatial norm $F(B_r)$ given by equation (3.6) as a function of time.

Despite our comments above, we would like to stress that *gufm1* does represent an excellent representation of the secular variation over the last four centuries. As an example, in figure 11 we show the fit of the model to the declination and inclination in London taken from the compilation of Malin & Bullard (1981). This data series has not contributed to the model; instead we use it as an external quality check. We do not use the data in the modelling because of the problem of error correlation due to the presence of the magnetic crust. For observatories that have measured the linear elements X, Y and Z, it is possible to account for the crustal anomaly, or bias, since it adds linearly to the field at any one time (see § 4 *c*). The London data consist of direction measurements only, so it is not possible to solve for biases in the same way, hence our omission of the data. Recall that the use of different observation sites is expected to generate a little over half a degree of noise in the observations. The overall misfit of the model to the 26 731 observatory annual means used in its construction is 1.21, slightly better than the misfit of *ufm1*, which is 1.32. These figures are calculated using the *a priori* errors estimates for each observatory described in Bloxham & Jackson (1992).

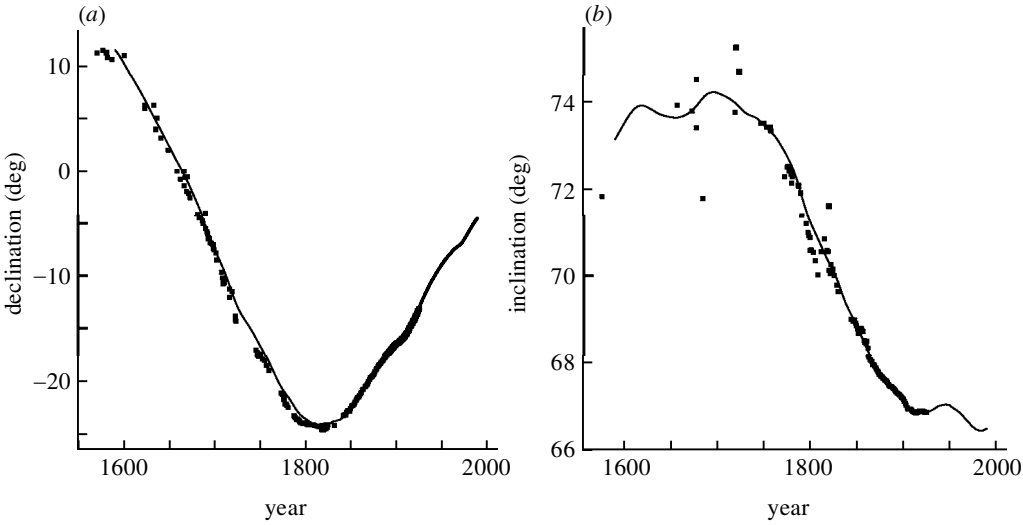

Figure 11. Fit of the model to the dataset of declination and inclination measurements compiled by Malin & Bullard (1981): (*a*) declination; and (*b*) inclination.

6. Discussion

It seems increasingly likely from palaeomagnetic studies that the mantle influences the field-generation process. Analyses of both sedimentary (Laj *et al*. 1991; Clement 1991) and volcanic (Love 1998) recordings of the last reversals suggest geographical biasing in the reversal paths of virtual geomagnetic poles (VGPs), an effect that would be predicted to be absent in the presence of homogeneous boundary conditions. The interpretation of the data is not incontrovertible however (e.g. Prévot & Camps, 1993; Merrill & McFadden 1999). Similarly, when examining the long-term stable magnetic field over the last few million years, persistent features can be found in some authors' models (Gubbins & Kelly 1993; Johnson & Constable 1995), although, again, the results are refuted by other authors (McElhinney *et al*. 1996; Carlut & Courtillot 1998).

The problem of deducing the pattern of flow at the surface of the core from geomagnetic observations is one that demands a high-quality model of the magnetic field in the past. Under the frozen flux approximation (Roberts & Scott 1965), all secular variation is ascribed to rearrangement of flux by surface motions; no diffusion of field is allowed. Recent work by Rau *et al*. (2000) has assessed this idea by examining its application to the output of numerical dynamo simulations. The study indicates that the approximation, in conjunction with the hypothesis of tangential geostrophy, works well, with the proviso that the calculation works at the scales appropriate to the underlying dynamo. It is imperative that the secular variation is known as accurately as possible, and, in our view, models that are based primarily on observatory data, such as the secular variation models of the IGRF, can be improved by a continuous representation of the field, such as that used both here and in constructing *ufm1* and *ufm2* of Bloxham & Jackson (1992), which allows all types of data to contribute to the estimate of field and its different rates of change. Models *ufm1* and *ufm2* have been used to investigate coupling mechanisms that could be responsible for changes in the length of day, and the model described here is equally suited for such studies.

The goals that spurred the data collection exercise on which we embarked in the mid 1990s were threefold.

(i) Firstly, we wanted to quantify properly the errors present in historical magnetic data more fully through a thorough analysis.

(ii) Secondly, we wanted to exploit the vast amounts of data that remain available in national libraries to improve the resolution of the core field back in time.

(iii) Thirdly, our aim was to develop a core field model that was suitable for studies of the flow at the top of the core, in particular for looking at large excitation events such as that which occurred at the end of the 19th century.

The work described here demonstrates that these goals have been met. A major disappointment has been the tiny amount of data that we found in the Spanish archives; we had hope of finding Pacific voyages to improve early coverage in this area. Our proper quantification of the observational errors in the data, as well as the determination of the accuracy of navigation in the pre-chronometer era, has not previously been described in the literature. Both of these effects determine the accuracy of the magnetic data, and, in taking account of the navigational errors, we have developed a new formalism to account for the error correlation that this effect is likely to generate.

Our model of the field at the core surface represents the field for the 400 year span 1590–1990. It improves on the models of Bloxham & Jackson (1992) by extending the time-span represented, and by being a single solution it removes any artefacts from the point at which the two models *ufm1* and *ufm2* join (1840). Our future work will centre around the testing of physical hypotheses regarding the field at the core surface, and the derivation of core motions from the database.

We have benefited from the help of a large number of colleagues in the course of this work. We are particularly indebted to Anne Murray for her fastidious work in various archives. Mioara Alexandrescu, Jose Luque and Stuart Humber have all helped with data collection, verification and homogenization. We thank Nick Barber for his development of the voyage editor software. David Barraclough, Ken Hutcheson and Jeremy Bloxham have shared their data and expertise with us, for which we are grateful. We thank Gary Egbert for supplying references to the Brownian motion literature and Jeff Love, David Gubbins, Cathy Constable and George Helfrich for useful discussions regarding the problem of navigational errors. This work was supported by NERC grants GR9/01848, GR3/10581 and The Royal Society; A.R.T.J. is supported by the Foundation for Historical Sciences with financial aid from The Netherlands Organisation for Scientific Research (NWO).

References

Andrewes, W. J. H. (ed.) 1996 *The quest for longitude*. Cambridge, MA: Harvard University Collection of Historical Scientific Instruments.

Backus, G. E. 1988 Bayesian inference in geomagnetism. *Geophys. J.* **92**, 125–142.

Backus, G. E. 1996 Trimming and procrastination as inversion techniques. *Phys. Earth Planet. Interiors* **98**, 101–142.

Backus, G. E. & Gilbert, F. 1970 Uniqueness in the inversion of inaccurate gross Earth data. *Phil. Trans. R. Soc. Lond.* A **266**, 123–192.

Backus, G. E., Parker, R. L. & Constable, C. G. 1996 *Foundations of geomagnetism*. Cambridge University Press.

Barraclough, D. R. 1974 Spherical harmonic models of the geomagnetic field for eight epochs between 1600 and 1910. *Geophys. J. R. Astr. Soc.* **36**, 497–513.

Barraclough, D. R. 1982 Historical observations of the geomagnetic field. *Phil. Trans. R. Soc. Lond.* A **306**, 71–78.

Bhattacharya, R. N. & Waymire, E. C. 1990 *Stochastic processes with applications*. Wiley.

Bloxham, J. 1985 Geomagnetic secular variation. PhD thesis, University of Cambridge.

Bloxham, J. 1986 Models of the magnetic field at the core–mantle boundary for 1715, 1777 and 1842. *J. Geophys. Res.* **91**, 13 954–13 966.

Bloxham, J. & Gubbins, D. 1985 The secular variation of Earth's magnetic field. *Nature* **317**, 777–781.

Bloxham, J. & Jackson, A. 1992 Time-dependent mapping of the magnetic field at the core–mantle boundary, *J. Geophys. Res.* **97**, 19 537–19 563.

Bloxham, J., Gubbins, D. & Jackson, A. 1989 Geomagnetic secular variation. *Phil. Trans. R. Soc. Lond.* A **329**, 415–502.

Carlut, J. & Courtillot, V. 1998 How complex is the time-averaged geomagnetic field over the past 5 Myr? *Geophys. J. Int.* **134**, 527–544.

Clement, B. 1991 Geographical distribution of transitional VGPs: evidence for non-zonal equatorial symmetry during the Matuyama–Brunhes geomagnetic reversal. *Earth Planet. Sci. Lett.* **104**, 48–58.

Constable, C., Parker, R. & Stark, P. B. 1993 Geomagnetic field models incorporating frozen flux constraints. *Geophys. J. Int.* **113**, 419–433.

Gubbins, D. 1975 Can the Earth's magnetic field be sustained by core oscillations? *Geophys. Res. Lett.* **2**, 409–412.

Gubbins, D. 1983 Geomagnetic field analysis. I. Stochastic inversion. *Geophys. J. R. Astr. Soc.* **73**, 641–652.

Gubbins, D. & Bloxham, J. 1985 Geomagnetic field analysis. III. Magnetic fields on the core–mantle boundary. *Geophys. J. R. Astr. Soc.* **80**, 695–713.

Gubbins, D. & Kelly, P. 1993 Persistent patterns in the geomagnetic field during the last 2–5 Myr. *Nature* **365**, 829–832.

Hansteen, C. 1819 *Untersuchungen über den Magnetismus der Erde*. Christiania, Norway: Lehmann & Gröndahl.

Holme, R. & Jackson, A. 1997 The cause and treatment of anisotropic errors in geomagnetism. *Phys. Earth Planet. Interiors* **103**, 375–388.

Hulot, G., Khokhlov, A. & LeMouël, J.-L. 1997 Uniqueness of mainly dipolar magnetic fields recovered from directional data. *Geophys. J. Int.* **129**, 347–354.

Hutcheson, K. A. & Gubbins, D. 1990 Earth's magnetic field in the seventeenth century. *J. Geophys. Res.* **95**, 10 769–10 781.

Jackson, A. 1989 The Earth's magnetic field at the core–mantle boundary. PhD thesis, University of Cambridge.

Jackson, A. 1990 Accounting for crustal magnetization in models of the core magnetic field. *Geophys. J. Int.* **103**, 657–673.

Jackson, A. 1994 Statistical treatment of crustal magnetization. *Geophys. J. Int.* **119**, 991–998.

Jackson, A. 1996 Bounding the long-wavelength crustal magnetic field. *Phys. Earth Planet. Interiors* **98**, 283–302.

Johnson, C. L. & Constable, C. G. 1995 The time-averaged geomagnetic field as recorded by lava flows over the past 5 Myr. *Geophys. J. Int.* **122**, 489–519.

Johnson, C. L. & Constable, C. G. 1997 The time-averaged geomagnetic field: global and regional biases for 0–5 Ma. *Geophys. J. Int.* **131**, 643–666.

Jonkers, A. R. T. 2000 North by Northwest: seafaring, science and the Earth's magnetic field. PhD thesis, Vrije Universiteit, Amsterdam.

Kellogg, O. D. 1929 *Foundations of potential theory*. Springer.

Kircher, A. 1641 *Magnes sive de arte magnetica*. Coloniae Agrippinae.

Laj, C., Mazaud, A., Weeks, R., Fuller, M. & Herrero-Bervera, E. 1991 Geomagnetic reversal paths. *Nature* **351**, 447.

Lancaster, P. & Salkauskas, K. 1986 *Curve and surface fitting: an introduction*. Academic.

Langel, R. A. 1987 The main field. In *Geomagnetism* (ed. J. A. Jacobs), vol. I, pp. 249–512. Academic.

Love, J. J. 1998 Paleomagnetic volcanic data and geometric regularity of reversals and excursions. *J. Geophys. Res.* **103**, 12 435–12 452.

Love, J. J. & Gubbins, D. 1996 Optimized kinematic dynamos. *Geophys. J. Int.* **124**, 787–800.

Ma, S.-Z. 1998 Spherical isotropic stochastic vector field in geomagnetism. *Chinese J. Geophys.* **40**, 575–585.

McElhinney, M. W., McFadden, P. L. & Merrill, R. T. 1996 The time-averaged paleomagnetic field. *J. Geophys. Res.* **101**, 25 007–25 028.

Malin, S. R. C. & Bullard, E. C. 1981 The direction of the Earth's magnetic field at London, 1570–1975. *Phil. Trans. R. Soc. Lond.* A **299**, 357–423.

Merrill, R. T. & McFadden, P. L. 1999 Geomagnetic polarity transitions. *Rev. Geophys.* **37**, 201–226.

Mountaine, W. & Dodson, J. 1757 A letter to the Right Honourable the Earl of Macclesfield, President of the Council and Fellow of the Royal Society, concerning the variation of the magnetic needle; with a set of tables annexed, which exhibit the results of upwards of fifty thousand observations, in six periodic reviews, from the year 1700 to the year 1756, both inclusive; and are adapted to every 5 degrees of latitude and longitude in the more frequented oceans. *Phil. Trans. R. Soc. Lond.* **50**, 329–350.

O'Brien, M. S. & Parker, R. L. 1994 Regularized geomagnetic field modelling using monopoles. *Geophys. J. Int.* **118**, 566–578.

Prévot, M. & Camps, P. 1993 Absence of preferred longitude sectors for poles from volcanic records of geomagnetic reversals. *Nature* **366**, 53–57.

Rau, S., Christensen, U., Jackson, A. & Wicht, J. 2000 Core flow inversion tested with numerical dynamo models. *Geophys. J. Int.* (In the press.)

Roberts, P. H. & Scott, S. 1965 On the analysis of the secular variation. I. A hydrodynamic constraint: theory. *J. Geomagn. Geoelect.* **17**, 137–151.

Rygaard-Hjalsted, C., Constable, C. G. & Parker, R. L. 1997 The influence of correlated crustal signals in modelling the main geomagnetic field. *Geophys. J. Int.* **130**, 717–726.

Sabine, E. 1868 Contributions to terrestrial magnetism. No. XI. *Phil. Trans. R. Soc. Lond.* A **158**, 371–416.

Sabine, E. 1872 Contributions to terrestrial magnetism. No. XIII. *Phil. Trans. R. Soc. Lond.* A **162**, 353–433.

Sabine, E. 1875 Contributions to terrestrial magnetism. No. XIV. *Phil. Trans. R. Soc. Lond.* A **165**, 161–203.

Sabine, E. 1877 Contributions to terrestrial magnetism. No. XV. *Phil. Trans. R. Soc. Lond.* A **167**, 461–508.

Shure, L., Parker, R. L. & Backus, G. E. 1982 Harmonic splines for geomagnetic field modelling. *Phys. Earth Planet. Interiors* **28**, 215–229.

Stevin, S. 1599 *De havenvinding*. Leiden: Plantiin.

Van Bemmelen, W. 1899 *Die Abweichung der Magnetnadel: Beobachtungen, Säcular-Variation, Wert- und Isogonensysteme bis zur Mitte des XVIIIten Jahrhunderts*. Supplement to *Obsns R. Mag. Met. Observatory, Batavia* **21**, 1–109.

Van Huffel, S. & Vandewalle, J. 1991 *The total least squares problem: computational aspects and analysis*. SIAM.

Wessel, P. & Smith, W. H. F. 1991 Free software helps map and display data. *Eos* **72**, 445–446.

Whaler, K. A. & Gubbins, D. 1981 Spherical harmonic analysis of the geomagnetic field: an example of a linear inverse problem. *Geophys. J. R. Astr. Soc.* **65**, 645–693.

Wright, E. 1657 *Certaine errors in navigation detected and corrected*. London: Joseph Moxton.